FEBRUARY 2024

Integrating Nonmarket Consumption into the Bureau of Labor Statistics Consumer Expenditure Survey

Ajit Zacharias, Fernando Rios-Avila, Nancy Folbre, and Thomas Masterson

A Study Conducted in Support of the BLS Research-Based Consumption Measure

Prepared under contract with the Bureau of Labor Statistics by the Levy Economics Institute of Bard College
(9/30/2021–2/27/2023, Order No. 1605C5-21-P-00021)

2/22/2024

BLS

Levy Economics Institute of Bard College

Integrating Nonmarket Consumption into the BLS Consumer Expenditure Survey

by Ajit Zacharias, Fernando Rios-Avila, Nancy Folbre, and Thomas Masterson

A Study Conducted in Support of the BLS Research-Based Consumption Measure

Prepared by the Levy Economics Institute of Bard College (Order No. 1605C5-21-P-00021)

February 22, 2024

Acknowledgements: We are grateful to the US Bureau of Labor Statistics for assigning us this project that enhances our research in the Institute's programs on Distribution of Income and Wealth and Gender Equality and the Economy. We also appreciate the valuable and insightful feedback we received from the BLS economists. In particular, we wish to thank Thesia I. Garner, Jake Schild, and Harley Frazis. We also wish to thank Lindsey Carter for her excellent editing assistance and Michael Stephens for managing the administrative aspects of the project.

Contents

List of Tables

List of Figures

Executive Summary

Consumption expenditures are a powerful indicator of standard of living because most families acquire key requirements for their daily lives—food, clothing, cellphone services, etc.—as commodities, i.e., goods and services exchanged for money. Economists continue to devote considerable effort to understanding how the acquisition of commodities translates into actual consumption and how living standards also depend on products that are not commodities. An important category of such products that are crucial to sustaining living standards results from household production activities and consist primarily of nonmarket and nongovernmental services such as do-it-yourself home repairs and childcare. Consumer Expenditure Surveys (CE) conducted by the BLS do not collect information on home production activities. The lack of data is the chief obstacle to developing household-level consumption measures that include household production. To fill this gap, the Bureau of Labor Statistics (BLS) awarded a contract to the Levy Institute in late 2021. This report contributes to the efforts of the BLS by creating a methodology to integrate information on household production. The quantities of the services resulting from household production are proxied by the labor inputs into their provision, i.e., the time allocated to providing the services. We include in the latter the time spent by household members for their own household. We also include the time allocated to services received by household members by people outside their household, e.g., care for older household adults with frailties provided by relatives or friends. Household production is defined as encompassing both sets of activities. The activities comprising household production contribute to household consumption; hence, some measure of its monetary value should be included in a broad measure of consumption.

We propose data sources and methods to build home production estimates in the CE Interview and Diary samples at the consumer-unit level. Our procedures are deployed on the quarterly CE public-use files that contain expenditure information for the calendar year 2019. The dataset we construct allows joint examination of consumption expenditures and household production by adding variables associated with the latter to each CE sample (Interview and Diary).

The 2019 round of the American Time Use Survey (ATUS), also carried out by the BLS, is our data source for the services provided by household members for their own households. Using 24-hour time diaries, the ATUS captures how *one* respondent (15 years or older) in a sample household spent a day and classifies the activities into detailed groups. Given our schema (described below) of constructing monetary values of household production, we categorize the activities into five groups: cooking, including clean-up (e.g., preparation of meals, washing dishes); housework other than cooking (e.g., doing laundry, cleaning the living premises, shopping for groceries, etc.); caring for and helping household adults (e.g., giving medicine, helping with a cellphone, etc.); active childcare; and supervisory child care. Standard definitions of childcare include only active care (e.g., reading to a child, bathing a child, etc.). However, we also account for passive or supervisory care of children (e.g., keeping an eye on a child while watching TV). The latter is not an activity. Instead, it measures the time household children were in the respondent's care. We differentiate between supervisory childcare that overlaps with cooking, other housework, or adult care and supervisory childcare that does not coincide with those activities. Supervisory care is intrinsically vital as care. Further, it can often constrain individuals (especially women) from engaging in other activities, particularly employment. In light of these considerations, we recommend that supervisory care be included in the definition of household production.

We use two sources of data to measure care services received by households from people outside their households. The people outside the household providing nonmarket care services may be family members who live outside the household (e.g., siblings, step or biological parents or children), members of the extended family (e.g., grandmothers or nieces), or nonrelatives (e.g., friends or neighbors). One of our data sources—the 2019 round of the Early Childhood Program Participation Survey (ECPP) module of the National Household Education Survey (NHES)—provides information to estimate weekly hours of nonmarket and nongovernmental care received by children under six. The other data source was the 2016 round of the Health and Retirement Survey (HRS) conducted by the University of Michigan. The HRS collects information on hours of assistance on certain activities received from people outside the household by older adults (51+ years) who report some health difficulty or disability.

Both sources would understate the care received. Several children older than six are also likely to be care recipients for childcare. Similarly, we ignore the external care received by people younger than 50 years of age (especially the younger disabled population) and care related to activities other than those captured in the HRS. Despite the limitations, we used these data sources because they appear to be the best currently available data for our purposes. We recommend that efforts be enhanced to gather information on nonmarket care received from nonhousehold members by older children (5 to 13 years) and disabled or sick individuals under 50 years of age living in households.

Estimates of home production that we build in the CE samples are based on imputation techniques widely used in the literature. A central statistical assumption involved in imputation is the Conditional Independence Assumption (CIA). Simply put, the premise is that, once we control for a set of relevant characteristics, the distribution of variables that cannot be observed jointly (in our case, consumption expenditures and household production) and require imputation is as good as independent. We tested this assumption using the 2019 data on time use and expenditures from the Panel Study of Income Dynamics (PSID) conducted by the Institute for Social Research, University of Michigan. The results suggest that the CIA is valid, which is a reasonable source of confidence in the imputations we carried out. At the same time, we should also point out that the definitions and measurement of consumption expenditures and time use are quite different between the PSID on the one hand, and the CE and ATUS on the other. Further, other assumptions (the validities of which are harder to assess than the CIA using available data, or even impossible) are involved in the estimates of home production we construct. On balance, our stance is that unknowable biases from assumptions that are difficult to assess empirically should not deter us from imputations necessary for understanding critical economic phenomena.

We implemented three methods of imputation from the ATUS. The simplest method we employed may be the "regression-prediction" (RP) method. Multivariate models were first estimated separately for individuals in households with and without children in the ATUS. The dependent variables in each model were the time allocated (weekly hours) to the five categories of household production in the ATUS, and independent variables consisted of a set of relevant individual and household characteristics. In the next step, we used the estimated coefficients to predict the hours of each person (15 or older) in the CE samples.

Our second imputation method was the technique known as multiple imputation (MI). Unlike the RP method, the MI procedure produces several imputed values of the same variable for each record (e.g., five different values are generated of the time allocated to cooking for each person). Admittedly, imputation errors can be accounted for coherently when using the imputed data for statistical inference (e.g., constructing confidence intervals for the average hours allocated to cooking). The multiple values of the same variable for each person are generated via a simulation using the variance–covariance matrix of the residuals of the estimated multivariate model (similar to that used in the RP method described above) in the ATUS. In light of the current software capabilities, we tested two models for this exercise—OLS and interval regression—and found both to produce similar results, although the latter is methodologically superior.

The third method utilized in the ATUS data was statistical matching (SM). Our matching algorithm has two crucial features: constrained matching based on propensity scores and stratification. The constraint is that matching is done without replacement (i.e., all the records in the donor file are used) to equalize the weighted number of observations in the ATUS ("donor file") and each CE sample ("recipient file"). Once the transformation required to satisfy the constraint is implemented (along with a few other steps), the following procedure is undertaken separately for each recipient file (i.e., CE Interview and Diary samples). We stack the donor and recipient files to form a single file. Then, we estimate propensity scores via a logit model with the probability of a record belonging to either the recipient or donor file as a dependent variable and a set of individual and family characteristics as independent variables. To further improve matching quality, stratification is carried out to form relatively homogenous groups in both files.

Small subgroups (e.g., subgroups formed by gender and the presence of children) are created and identified using principal component analysis (PCA) and cluster analysis. We iteratively match observations based on rank similarity within subgroups, modifying the subgroups as needed in the process to ensure that one of all the possible candidates from the donor file is selected as the effective donor for each record in the recipient file.

For imputations based on the ECPP and HRS data of services for children and older adults received from people outside the home, we utilized a two-stage imputation procedure. First, the likelihood of receiving care is modeled via a binomial logit model. This model predicted the propensity score of recipiency in the CE samples. A stochastic component was added to the prediction by assigning recipiency based on a draw from a Bernoulli distribution—using the predicted score for each individual as a parameter for the draw. In the next step, we estimated a Poisson model of the hours of care received using only the subsample that received care. The model parameters were then employed to estimate hours of care received by those assigned recipiency of care in the CE samples. We assessed the quality of imputations by assessing how close the incidence of the recipiency of care and the average hours of care—computed from the imputed data—were to those estimated from the original data. In general, we found that the estimates were satisfactorily close. A more detailed examination of the imputation quality was not conducted, nor were alternative imputed values created because a preliminary perusal of the data showed that the incidence and quantum of external care was small relative to “internal” household production, i.e., services received by the household from members of the same household.

In contrast, our final data product provides imputed values produced by the three alternative methods from the ATUS. They were also subjected to much closer scrutiny. We recognize that methodological preferences and analysis goals would differ among the potential users of the imputed data. To provide flexibility and room for sensitivity tests, we recommend that imputed values generated by all methods be made available for public use. We found that all three methods performed equally well in replicating the average values of time allocated by individuals in various demographic subgroups, except for some male subgroups where the RP and MI methods did not perform as well as SM. The advantage of SM over the other two methods was clearly evident in replicating the median values and other percentile points in the distribution of time allocated (e.g., the 75th percentile value of the time given to cooking by employed women). On the other hand, advanced users interested in multivariate analysis of imputed data may favor the MI method because of the well-established statistical inference techniques associated with the method available in statistical software such as SAS®.

Once the imputations described above are carried out for each individual in the CE samples, we construct the aggregate for the consumer unit by adding up the imputed amounts for individuals in the consumer unit. For example, the time allocated to active childcare by the consumer unit is obtained by summing the time allotted by each individual in the consumer unit to active childcare. To avoid double counting in calculating the total time spent on household production by each individual, the time during which supervisory care coincides with other household production was included only in *one* of the following categories: cooking, housework other than cooking (including shopping), or adult care. Supervisory care that does not overlap with household production was obtained by subtracting overlapping supervisory care time from the total time allocated to supervisory care.

Further, supervisory childcare time in consumer units with multiple adults was adjusted to avoid the potential double-counting of such care (e.g., both parents may keep an eye on their toddler while watching TV during a given time of the day). The adjustments were made on parameters differentiated by gender and the number of likely caregivers estimated from the ATUS. Better information is required to understand the dynamics of supervisory care provision and other domestic services in households with multiple providers and receivers. Unfortunately, this cannot be obtained under the current ATUS sampling strategy of collecting time diaries from only one respondent per household.

The final step in our estimation endeavor was valuation, i.e., the conversion of hours of household production into monetary values. Given the alternative perspectives on the valuation principle, we believe that at least two sets of monetary values should be provided in the public-use file. We produced one set of values using the so-called generalist approach that applies a single hourly wage rate to various tasks of household production. Perhaps the most well-known estimate embodying this principle is the value of household production calculated by the US Bureau of Economic Analysis as a supplement to national accounts. Our preferred method is often described as the specialist-wage approach because it uses different wage rates to value various tasks of household production. Specifically, we used the average hourly wage for cooks, workers in the "private household" industry, preschool and kindergarten teachers, and nursing assistants to value the time allocated to cooking, housework other than cooking, active childcare, and adult care. Because supervisory childcare is less costly to replace than active childcare, we used half of the average wage of preschool and kindergarten teachers to assign a monetary value to supervisory childcare. For childcare from people outside the household, we apply the same rate as active childcare rendered by a household member. Similarly, we value care for adults given by people outside the household at the same rate as members of the home. After completing the valuation of the various categories according to each principle, the resulting monetary values were aggregated to form the value of household production for each consumer unit in the CE samples.

While the central goal of our research was to develop a methodology for integrating household production into the CE samples, we conducted a preliminary descriptive analysis of the imputed values. Some tentative conclusions emerged. Irrespective of the valuation principle, the impact of adding the value of household production to consumption expenditures[1] is enormous: the average value of expenditures that includes household production ("augmented expenditures") is higher than a standard definition by 50 and 63 percent for the generalist and specialist principles of valuation, respectively. The provision of these services—which adds to potential consumption but is excluded in the standard measures—is primarily taken on by women, reflecting the patriarchal allocation of household production responsibilities in our society. We estimate that women accounted for an overwhelming 78 percent of the total value and hours of household production in 2019.

The measured disparities in consumption expenditures between population subgroups are affected by the inclusion of household production. Crucial demographic variables at work are the number of adults and children in the household. Ignoring home production leads us to conclude that the highest average expenditures are among households with two adults and two children. However, including home production puts the group with three or more adults and three or more children at the top position. Incorporation of household production also causes a re-ranking among households differentiated by the race or ethnicity of the consumer unit's reference person. We found that the change in racial disparities was primarily linked to differentials among groups in the incidence of households with children. The most striking finding in this regard was the reversal of rank between Hispanic and White families: Hispanics ranked slightly above Whites regarding average expenditures when we added household production to spending.

Consistent with the spirit of the earlier studies, we also found that household production has an inequality-reducing effect on the distribution of consumption expenditures. Our estimates show that for consumer units in the bottom half of the standard expenditure distribution, the average value of household production is approximately equal to average expenditures. For those in the top half, the average value of household production shows only a small variation across expenditure deciles compared to the sharp gradient of average expenditures. Thus, the relatively large size of household production and its somewhat equal distribution across the expenditure distribution ensures that its addition will increase the expenditure shares (e.g., the share of the bottom decile in the aggregate value of the new expenditures that

[1] Our starting point in estimating consumption expenditures is the total consumer expenditures available in the public use CE data. The latter consists of expenditures in 14 major categories: food, alcohol, housing, clothing, transportation, healthcare, entertainment, personal care, reading, education, tobacco, miscellaneous, cash contributions, and personal insurance and pensions. We subtract from the BLS definition of consumer expenditures, mortgage interest, property taxes, maintenance, repairs, insurance, and other expenditures on owner-occupied primary home. We also subtract expenses classified as miscellaneous, cash contributions, and personal insurance and pensions. Finally, we add the rental equivalent for owned homes.

includes the value of household production) of those in the lower rungs of the distribution. This shift will be reflected in the new distribution, displaying a lower inequality than the prior distribution.

Two opposing forces affect the change in inequality. First, since the mean value of household production is notably higher for those with children than for those without, adding home production to expenditures widens the intergroup inequality in the augmented measure compared to the standard estimate of expenditures. However, because of the lesser inequality in the distribution of household production than in that of spending, the inequality in the augmented measure within both groups will be lower than in the standard measure. We found, via a decomposition analysis, that the reduction in within-group inequality overwhelms the increase in between-group inequality. Additionally, the inclusion of supervisory care in the definition of childcare has an inequality-enhancing effect compared to a measure of childcare that includes only direct childcare. As noted above, the value of supervisory care is roughly equivalent to that of direct childcare. Hence, including supervisory care widens the between-group inequality and increases overall inequality.

In sum, our research has provided an empirical methodology and has identified data sources to extend the consumer expenditure data collected by the US Bureau of Labor Statistics by incorporating household production. We applied and tested our approach by generating estimates for 2019. Our recognition of the diversity of views and analytical strategies among researchers studying household production and its relationships with living standards has led us to implement estimation of various categories and imputation techniques to accommodate at least some of the diversity. Instead of a single homogeneous measure of household production, we also generate estimates of the major components of household production, such as childcare and cooking. This feature permits, for example, the analysis of the relationship between expenditures on prepared food and time allocated to cooking. We also provide imputed values for the principal categories of household production (except care services received from persons outside the household) based on three alternative imputation methods. Therefore, researchers can examine the sensitivity of their findings to alternative methods of imputation and use a particular set of imputed values based on their analytical strategy and goals. In a similar vein, we also generate monetary values of household production and its subcategories generated according to two principles of valuation. We hope that our efforts contribute to closer scrutiny of household production and aid in the formulation of policies to reduce gender inequality and economic inequality.

Research Goals

The research conducted by the Levy Institute for the US Bureau of Labor Statistics (BLS) under Order No. 1605C5-21-P-00021 aimed to develop methods and identify data sources to build home production estimates in the CE Interview and Diary samples. Our research also involved applying the empirical methodology to the quarterly CE public-use files that contain expenditure information for the calendar year 2019. Specifically, we undertook the following activities, which are discussed in detail in the remainder of this report:

- Identify an appropriate household production definition to integrate with consumer expenditure data in light of existing research and practices.
- After reviewing alternative approaches, identify two alternative valuation methods for household production and the best available data to estimate the valuation rates (i.e., the rates at which household production hours of various types are to be converted into dollars).
- Provide a literature review of imputation of household production and assessment of imputation methods.
- Conduct a validation exercise examining the critical assumption of conditional independence using the Panel Study of Income Dynamics that contains joint information on household production and consumption expenditures.
- Implement imputation using statistical matching via matching records in the 2019 ATUS weekend and weekday diaries separately to individual records in each quarterly CE public-use file for each Interview and Diary sample.
- Perform using the method of regression-prediction, imputation of six categories of time allocated to household production from the 2019 ATUS weekend and weekday diaries separately to individual records in each quarterly CE public-use file for each Interview and Diary sample.
- Implement employing the method of multiple imputation, imputation of six categories of time allocated to household production from the 2019 ATUS weekend diaries and weekday diaries separately to individual records in each quarterly CE public-use file for each Interview and Diary sample.
- Impute hours of childcare received from people outside the household based on the 2019 round of the Early Childhood Program Participation Survey (ECPP) to each person under six years in each quarterly CE public-use file for each Interview and Diary sample.
- Impute hours of care of older adults received from people outside the household based on the 2019 round of the Health and Retirement Survey (HRS) to each person over 50 years in each quarterly CE public-use file for each Interview and Diary sample.

- Conduct an extensive quality assessment of the imputations.
- Combine the results of imputations at the individual-level to generate consumer-unit level estimates of household production hours in the CE samples.
- Use the consumer-unit level estimates of household production hours to generate the monetary values of household production according to two valuation principles.
- Provide a preliminary analysis of the joint distribution of household production and consumption expenditures using descriptive statistics and decomposition analysis of the third-quarter CE Interview file.
- Explore (tentatively) the relationship between expenditures of time and money for food and childcare separately using the third-quarter CE Interview file.
- Provide recommendations to the BLS to aid their planned work on home production. The recommendations pertain to the scope and valuation of household production, imputations of various categories of household production, and use of CE data with imputed home production variables.

1 Introduction

The United States is among the handful of industrialized nations to publish official statistics on nonmarket production as a part of its national accounts. It has also been at the forefront of collecting annual data on time use for the last two decades via the American Time Use Survey (ATUS) of the US Bureau of Labor Statistics (BLS)—the vital input required for the estimates on nonmarket services in the national accounts. Recently, the BLS launched a laudable initiative to explore the integration of home production into the Consumer Expenditure Survey (CE) as part of an effort to develop a broad measure of household economic well-being (Armstrong et al. 2022). As a part of the initiative, the Bureau of Labor Statistics (BLS) awarded a contract to the Levy Institute in late 2021. Accordingly, we propose methods and data sources to build home production estimates in the CE Interview and Diary samples at the consumer-unit level. The dataset we construct allows joint analysis of consumption expenditures and household production by adding variables associated with the latter to the CE samples.

Our starting point details a set of considerations in defining the scope of home production suitable for integration with a measure of consumption (Chapter 2). We argue that the definition of childcare should be expanded to include supervisory childcare—which is not an activity per se when performed on a nonmarket basis but a constraint, especially on women. In addition, we should consider nonmarket care received by members of the consumer unit from people outside their household (e.g., a grandmother caring for her grandchild while the parents of the child are at their jobs) because such services expand consumption possibilities. In contrast to the US national accounts approach of using a single hourly rate to convert the time spent on diverse home production activities into monetary units, we propose a set of specialist hourly rates. In particular, we assign a higher value to the active care of dependents than to routine housework. Our valuation proposal is motivated by the observation that market substitutes for direct care work are generally costlier than substitutes for routine housework under the existing occupational wage structure. Hence, they may reflect relevant market replacement costs better than the single hourly rate anchored to one of the occupations with relatively low wages, i.e., workers in private households.

Ideally, consumer-unit–level nonmarket consumption estimates should be built from information collected directly from the respondents in the CE samples. However, the CE does not collect such information. Consequently, reasonable guesses—"imputations"—have to be made regarding nonmarket provisioning undertaken by members of a consumer unit for their household ("self-provisioning") or nonmarket services received by the consumer unit from people outside their homes. Admittedly, no known imputation method will faithfully replicate the information that the individual respondent would have given if queried during the administration of the survey. Researchers have used various statistical methods for imputation, and we provide a brief overview in Chapter 3. In general, the best hope is to faithfully replicate the standard measures of location (mean and median values) and dispersion (standard deviations or measures of inequality such as the Gini coefficient) for a reasonable number of population subgroups (e.g., households with children and employed, female reference persons). In practice, the extent to which various methods would fulfill that hope differs.

Further complicating the issue is that there are no clear-cut scientific rules on which imputation performs best. Therefore, judgment calls and rules of thumb are required in this exercise. The main focus of Chapter 3 is to discuss how we used three different strategies to impute, using the 2019 ATUS file, self-provisioning undertaken by individuals aged 15 and over in the consumer units covered in the Interview and Diary samples for 2019. We also present results on the quality of the imputations. Additionally, Appendix A of the chapter reports results from a validation exercise using data from the Panel Study of Income Dynamics. The purpose of the exercise was to ascertain the validity of the conditional independence assumption—an assumption that is made in most methods of imputation. The results lend strong support to the assumption with regard to the relationship between total consumption expenditures and categories of household production. However, further research would be required to assess whether the same conclusion would apply to the relationship between subcategories of expenditures (e.g., food expenditures) and household production.

We also continue on the theme of imputations in the next chapter (Chapter 4). Here, the focus is on nonmarket care rendered by people outside the household to the youngest (under 6 years) and older (over 50 years) members of the consumer unit. In principle, we should account for the care received by members of all age groups. Unfortunately, reasonably reliable microdata is not available to facilitate such imputations. For the youngest group, we use the 2019 Early Childhood Program Participation (ECPP) survey that contains detailed information on the incidence and weekly hours of care received from relatives and nonrelatives, along with the demographic characteristics of the children and their families. Our source of information for the care received by older adults is the Health and Retirement Survey (HRS) of 2016. The HRS provides information on help received by adults with physical or mental frailties. We can identify the amount of care (in number of hours) received from people outside the household. When combined with data on the characteristics of the recipients and their households, the HRS can provide the ground for making reasonable imputations in the CE samples.

The imputations described in Chapters 3 and 4 allow us to accomplish the goal of producing estimates of home production (perhaps better described as nonmarket consumption in the current context) at the consumer-unit level in the CE Interview and Diary samples. Our final synthetic CE data files have estimates of home production by members of the consumer unit using three imputation methods.

Estimates in the files of help received from outside the household were generated with a single imputation method. Both categories of home production are valued by our preferred method ("specialist wages") as well as what we describe as the BEA (US Bureau of Economic Analysis) method, using the generalist wage. Results from alternative imputation methods are provided to facilitate the use of the data for various research purposes, from descriptive analysis to multivariate modeling. We also include results from two valuation methods for users to judge the sensitivity of their findings to the valuation method.

We use the synthetic Interview file for the third quarter of 2019 to present some salient features of the relationship between consumption expenditures and home production (Chapter 5). Irrespective of whether one takes supervisory childcare or care received from outside the household, and under both valuation methods, the inclusion of home production has a big impact. For the entire sample, average consumption that includes home production ("augmented expenditures") is higher than standard consumption expenditures by at least 50 percent. The effect on families with children is even larger. We also briefly discuss the impact of including home production on consumption inequality. Echoing the findings from similar studies that incorporate home production in an income measure, we find that the measured level of inequality falls significantly, regardless of the definition or valuation of household production. Once again, substantial differences emerge between the impacts on households with and without children. Further, these differences are notably affected by whether we include supervisory childcare. We hope our initial foray into using the synthetic files leads to further scrutiny of these and related issues.

In the final chapter (Chapter 6), we outline the study's recommendations for integrating home production into the CE. These recommendations are derived from the conceptual and practical considerations outlined in Chapters 2–4. We were also guided by the data analysis that we report here. In addition, we took into account the findings from several other explorations of alternative assumptions and techniques, which are not reported here because of space limitations. Finally, we provide a set of guidelines that may be helpful to users of the data.

2 Scope and Valuation of Household Production

2.1 Introduction

Economists have long recognized that household production is crucial in sustaining living standards by providing goods and services (see, e.g., Antonopoulos 2008; Folbre 2009; and Wolff and Zacharias 2007). The activities included in the household production orbit can generally be delegated by the household member performing that activity to someone outside the household for pay—the so-called "third party principle," devised by Margaret Reid (1934). Hence, leisure (or, for that matter, those activities related to formal learning) and personal care are excluded from the scope of household production. Discussions of various methods of putting a monetary value on household production in the US date back to Kneeland (1929) and Reid (1934). In his work on national income accounts, Kuznets (1946) acknowledged the principle of using a "housekeeper wage" method, usually described today as the "generalist wage" method. Countless generations of students have since heard about how GDP would decrease when the economist marries his or her housekeeper, illustrating the omissions in national accounts. There are several excellent overviews of valuation methods (e.g., Goldschmidt-Clermont 1993; National Research Council 2005). In light of the existing literature, our focus in this chapter is to outline our methodological and empirical choices in defining and valuing household production. Since official US economic statistics, usually dubbed "satellite accounts," already include aggregate estimates of the value of household production, we compare and contrast our procedure with the existing official methodology. We feel that such an exercise is particularly appropriate because the present research is expected to contribute to an expanded measure of consumption that another government agency, the Bureau of Labor Statistics, plans to assemble.

2.2 Definitions

We chose a household production definition suitable for a broad consumption measure at the household or consumer-unit level. Accordingly, our measurement excludes nonmarket services for people outside their household (e.g., friends or relatives) rendered by household members. By the same logic, we include nonmarket services received by members of the consumer unit from people outside their household because those services expand the consumption possibilities of the consumer unit—our object of measurement. In contrast, who receives the services is not relevant for aggregate measures of household production. Hence, the satellite account of household production developed by the US Bureau of

Economic Analysis (BEA) includes the time spent on providing nonmarket services to nonhousehold members in its scope (Bridgman, Craig, and Kanal 2022).

Household production can include the production of both goods and services. Historically, in the United States, home production of various food items and clothing was a substantial contribution made by women to sustain family living standards (Kneeland 1929).[2] In farm families, some of the output is consumed on the farm, not just for intermediate but also for final uses. The aggregate personal consumption expenditures in the official US national accounts include the imputed value of farm products destined for final consumption of the farm families (Line 202, Table 7.12).[3] In 2021, it amounted to roughly $200 million, a tiny percentage of personal consumption expenditures. Increasing commodification of everyday life, i.e., growth in the share of business enterprises in the supply of consumer goods, has seen the category of goods gradually disappear from production for own use by households. No systematic data exists regarding the output of goods for own use, such as nonfarm production of fruits and vegetables.[4]

A lacuna exists on the input side as well. However, the ATUS includes several categories of time-use, at least some of which may be dedicated to producing goods for own use. These categories are the following (with examples indicating an activity that leads to the production of a good provided in the parentheses after each type): “Sewing, repairing and maintaining textiles” (knitting sweaters), “Food and drink preparation” (brewing beer), “Building and repairing furniture” (building a table), “Heating and cooling” (chopping firewood), and “Lawn, garden and houseplant care” (planting herbs or vegetables). Additionally, the ATUS includes two categories that may include the time spent on producing tangible assets that provide consumption flows for several years: “Exterior repair, improvements, and decoration” (building a garage) and “Ponds, pools, and hot tubs” (e.g., putting in a pool). While we cannot explicitly identify the time spent on producing goods or tangible assets per se, we include these broader categories of time use in our definition of household production.

Turning to the self-provisioning of services, we can categorize the activities into housework (e.g., cleaning the home) and direct care of household members in need of assistance. Our preferred definition of childcare is broader than that employed in the BEA definition. Folbre et al. (2005) have argued that the standard childcare measure focuses on activities alone and can be considered active childcare. However,

---

[2] For a critical historical interrogation of the notion of the homemaker, see Coleman (1998); Shapiro (2009) provides a social history of the transformation of production of food at home.

[3] Table 7.12 “Imputations in the National Income and Product Accounts” (Last revised on September 30, 2022).

[4] The ATUS does collect information on the time spent on “lawn, garden and houseplant care” (020501), which should include the time inputs into growing fruits, vegetables, and herbs, and travel related to this activity. We have included this category in our definition of housework described below.

active childcare underestimates the time demands of childcare because it ignores the on-call aspect of care responsibilities. Suh and Folbre (2015) describe the latter as the time spent on supervisory care and propose a methodology to measure it in the ATUS, which we follow. Purchases of childcare services outside the home tend to reduce supervisory constraints far more than active childcare time, a finding supported by analysis of data from the Panel Study of Income Dynamics as well as analysis of the impact of maternal employment hours on active and supervisory care in the ATUS (Gautham and Folbre, forthcoming, Suh and Folbre 2023).

Definitions of time devoted to household production categories by members of the household for the household are operationalized using the codes from the ATUS activity lexicon pertaining to the 2019 round and, for supervisory care, are based on the measure of "in your care" time that the ATUS labels supervisory care, adjusted by some other variables (such as overlaps with other adults) available in the public-use data file (Table 2-1). Household production by household members for their household (self-provisioning) consists of five broad categories: cooking, housework (except cooking), shopping, care of household adults, and care of household children. Further, we list two categories of nonmarket care the household receives from people outside the home: care of children and adults. They draw on data sources other than the ATUS and are discussed in detail later (Chapter 5).

Table 2-1 Categories of household production included in the study

Number	Time category	ATUS codes, ATUS variable name, or Description of data source
1	**Cooking**	
	Food and drink preparation; presentation, and clean-up	0202
	Related travel	180202
2	**Housework (except cooking)**	
	Housework (except cooking)	02 (except 0202)
	Related travel	1802, excluding 180202
3	**Shopping**	
	Using financial services	0802
	Using legal services	0803
	Using household services	0901
	Using home maint/repair/décor/construction services	0902
	Using pet services	0903
	Using lawn and garden services	0904
	Using vehicle maintenance or repair services	0905
	Using government services	1001
	Telephone calls to/from salespeople	160104

Number	Time category	ATUS codes, ATUS variable name, or Description of data source
	Telephone calls to/from household services providers	160106
	Travel related to consumer purchases	1807
	Travel related to using financial services	180802
	Travel related to using legal services	180803
	Travel related to using household services	180901
	Travel related to using home maint/repair/décor/construction services	180902
	Travel related to using pet services	180903
	Travel related to using lawn and garden services	180904
	Travel related to using vehicle maintenance or repair services	180905
	Travel related to using government services	181001
4	**Care of household adults**	
	Activities related to the care of household adults	0304, 0305, 0399
	Related travel	180304, 180305
5	**Childcare**	
	Active childcare	
	Activities related to the care of household children	0301, 0303,
	Using paid childcare services	0801
	Related travel	180302, 180303, 180801
	Supervisory childcare (in-your-care)	TRTHH_LN*
	Overlapping with other household production	Supervisory childcare while engaged in activities numbered 1,2, 3, or 4 above
	Non-overlapping with other household production	Supervisory childcare *less* overlapping supervisory care
6	**Care received from people outside the household**	
	Care received by children	Early childhood program participation survey
	Care received by adults	Health and retirement survey

*Total time spent on secondary childcare for household children under 13 years. The variable is available in the ATUS public-use data files.

As described above, we have included the category of supervisory childcare or "in-your-care" time. We measure it as the total time spent on providing secondary childcare for household children under 13 years of age. The ATUS considers the respondent as engaged in secondary childcare based on several questions regarding whether the person had a child under 13 in their care during the activity, among other considerations. In particular, if the activity involves sleeping, providing active childcare, or takes place when the children are asleep, the time for that activity is excluded from the measure of secondary care. We measure overlapping care time as time when a household child under 13 years was in the care of the respondent while the latter was engaged in activities we define as cooking, housework (except cooking), shopping, or care of household adults. For example, if a 5-year-old was in the respondent's care during the 30 minutes the respondent spent on cooking, we will count those 30 minutes as 30 minutes of overlapping care time. The non-overlapping amount is then calculated as a residual by subtracting the overlapping time from the total time spent providing secondary care.

The ATUS collects information on one respondent aged 15 or older from each sample household. We use this information to impute time spent on home production by each individual above 14 years of age in the CE Interview and Diary samples. We define the total time spent on household production by a household member as the sum of the time spent on cooking, other housework, shopping, care of household adults, active care of household children, and non-overlapping supervisory care of household children. To avoid double counting, we exclude the time spent on overlapping supervisory care from the calculation of total time spent on childcare and household production.[5]

As noted above, we also impute the hours spent by people outside the household to provide care to individuals in the CE samples. To obtain the consumer-unit level estimates of hours spent by household members for self-provisioning and spent on household members by those outside the household, we add up the hours of each person in the unit. An exception is made, however, for time spent on supervisory care. The adjustment is required because it is not uncommon for another adult to be present in the same room when an adult reports that a child under 13 is in their care (see Chapter 3 for the adjustment). We regard only the presence of one adult as socially necessary to provide supervisory care.

2.3 Valuation

Most literature on converting household production time into monetary values uses generalist, specialist, or own wages. The BEA, for example, uses the average hourly compensation of private household employees. The BEA method reflects the principle behind the generalist wage—using a single hourly rate to value the diverse tasks included in household production. On the other hand, advocates of the specialist wage argue that different broad categories of household production require the application of different hourly rates. For example, the National Research Council (2005) proposed using a specialist wage method whereby the specialist hourly rate is multiplied by a factor (a number ranging from zero to one). The

[5] Suppose that a person spends an hour cooking that does not overlap with supervisory care and another hour on cooking that overlaps with supervisory care. Also, assume that she or he does not spend time on any other housework, active childcare, or adult care. We will calculate the total time spent on household production by the person as the total time spent cooking, i.e., two hours. We avoid double counting by not considering the hour she or he spent cooking while engaged in supervisory care as an hour of cooking *and* an hour of supervisory care.

factor is supposed to reflect the relative productivity of the household member who performs the task (National Research Council 2005, 70). The third method of own wage is derived from the notion of opportunity costs and uses an hourly rate that corresponds to the household member's actual or potential wage.

The generalist and specialist wage methods are attempting to answer the question of market replacement: i.e., how much would it cost to procure market substitutes for home production? Advocates of generalist wages argue that the appropriate replacement is hiring household help to perform all housework and care for dependents. In contrast, the assumption behind the specialist wage is that replacement would entail hiring workers to perform different tasks that would otherwise have to be carried out by household members. The own wage approach does not address the same question of replacement. Instead, the strategy focuses on the wage the household member could earn if devoting the same time to paid employment.

We are in favor of the specialist wage method. However, given that our valuation exercise uses imputed data, we propose a lower level of disaggregation of activities than in some previous estimates of the value of nonmarket household services (Suh and Folbre 2015). Of the time categories included in our definition of household production, we consider "Cooking" as a separate category for valuation because previous research reveals considerable substitutability between purchased and home-produced meals. Another rationale for this distinction is that cooking is a more regular daily activity than many other categories of unpaid work. Our hourly rate for cooking is the average wage of cooks in various occupational groups we calculated from the 2019 Occupational Employment and Wages Survey (OEWS), a survey of establishments conducted by the BLS.[6]

Next, we combine the time categories "Housework (except cooking)" and "Shopping" into a single category for valuation. A generalist wage is assigned as the hourly rate for this category. Finally, we use the same source of data as in the BEA estimates, viz., the Merged Outgoing Rotations Group (MORG) earnings data compiled from the monthly CPS (2017–19 pooled sample) and calculate the hourly earnings of workers employed in the industry "Private households" (industry code 9290).

Methodological considerations regarding the use of imputed data also prompt us to consider active childcare as a single category rather than distinguishing among physical, developmental, and managerial

[6] The occupational groups are: Cooks, fast food; Cooks, institutions and cafeteria; Cooks, restaurant; Cooks, short order; Cooks, all other; and Cooks, private household. OEWS occupational codes for the groups are in the listed order: 35-2011, 35-2012, 35-2013, 35-2014, 35-2015, and 35-2019.

care.[7] We believe a separate specialist wage should be used for the valuation of unpaid childcare, based not on the occupational wage of childcare workers but on that of early childhood educators. Specifically, we use the mean combined hourly wage for the occupational groups "Preschool teachers, except special education" (OEWS code 25-2011) and "Kindergarten teachers, except special education" (OEWS code 25-2012). Much concern has been expressed about the underpayment of childcare workers in the US. For instance, the Build Back Better initiative recently passed by the US Congress envisages early childhood educators to be paid on par with K–12 educators. In some countries, including France, childcare workers are paid approximately the same as elementary school teachers.

Some previous research has valued developmental childcare at a higher rate than other childcare activities (Suh and Folbre 2015). As mentioned, we cannot meaningfully disaggregate childcare activities in this imputation. Most parents incorporate developmental efforts into their interactions with small children, and they develop child-specific skills that enhance their productivity in many respects. Some basic level of parental care is critical to successful child socialization and capability development. Many parents have less education than is characteristic of preschool and kindergarten teachers, and the home environment lacks the extra stimulation provided by children of similar age. On the other hand, the primarily one-on-one interactions of childcare in a home environment are typically high quality. Small children often enjoy the undivided attention of a parent or other family member.

The cost of out-of-home childcare is a more relevant replacement cost for most families than hiring household help, and it is undoubtedly utilized by a larger number of families with children. However, most estimates of the cost of out-of-home childcare services are based on weekly rather than hourly rates and utilization hours are heterogeneous. An additional consideration is that the cost of out-of-home childcare understates the value of childcare because it is strongly influenced by the relatively low pay of childcare workers, as we mentioned earlier. For these reasons, the input-based specialist replacement cost estimate for early childhood educators probably remains the best choice.[8]

[7] While these disaggregated categories will not be imputed on the household level, it will be possible to apply a set of ATUS averages for the composition of active childcare to households in the merged data set to provide an approximation.

[8] A recent study on Austria argues that the wage information from digital platforms is superior to that derived from labor force surveys in some respects; the platforms report wages specific to tasks that would be performed in individual households and are often geographically detailed (Jokubauskaitė and Schneebaum 2021). Several digital platforms operating in the US provide data on wages advertised by job applicants. However, aggregate data does not appear to be available (or scrapable) at this time. More importantly, the characteristics of the self-selected samples are unclear, and almost certainly more upscale. A methodological issue that emerges from the consideration of data from digital platforms is the economies of scale in childcare: the differential in hourly cost between in-home nanny care and center-based care shrink with the number of children under care. This has implications for differences between an input-based valuation based on replacement cost and an output valuation based on purchase of services outside the home (for a detailed discussion of this issue, see Mullan 2010).

We suggest a lower replacement wage rate should be assigned to supervisory care time than active care. While it entails considerable responsibility, supervisory care is also relatively flexible and less demanding than housework or active childcare. In addition, providing supervisory care in one's home is generally easier to combine with other activities in one's home than elsewhere. A further consideration is that there is some ambiguity about its developmental value. For these reasons, we employ a rate that is one-half the rate for active childcare.

As we noted, supervisory care is often performed with other activities. We present some tabulations of parental supervisory care time in Table 2-2 below, based on pooled ATUS data from 2004 to 2019. For fathers and mothers, three-quarters of supervisory care occurs while undertaking activities other than active childcare. However, fathers tend to be relatively more engaged in leisure activities than mothers while constrained by supervisory care. For mothers, there is almost an even split between leisure on the one hand and unpaid housework (including cooking) and shopping on the other hand. Since we do not include leisure in our definition of household production, an hour of supervisory care that occurs while engaged in leisure is valued, as mentioned above, at 50 percent of the rate for active care. However, an hour of supervisory care simultaneous with, say, cooking is effectively valued at a rate higher than the rate applied to an hour of cooking alone or an hour of supervisory care concurrent with leisure.[9]

[9] Suppose that a person spends an hour cooking that does not overlap with supervisory care and an hour on cooking that overlaps with supervisory care. She does not spend time on any other housework, active childcare, or adult care. Our valuation methodology will assign a monetary value of $2w_k + 0.5w_c$ to her household production, where w_k and w_c are the hourly wage rates for cooking and active childcare. Now consider another person that spends two hours on cooking and engages in no other household production. The monetary value of her household production will be $2w_k$. Since $2w_k + 0.5w_c > 2w_k$, we are valuing the first person's two hours at a higher effective rate.

Table 2-2 Percent of parental supervisory (in-your-care) time overlaps with leisure and unpaid housework and procurement activities in households with children, no other adults present (ATUS 2004–2019)

	Any child under 13	Children 0-4, none older	Children 5-12, none younger
Fathers			
In-your-care time overlapping with unpaid housework and shopping	25%	25%	24%
In-your-care time, overlapping with leisure	51%	50%	52%
In-your-care time, overlapping with either leisure or unpaid housework and shopping	75%	75%	76%
Mothers			
In-your-care time, overlapping with unpaid housework and shopping	38%	37%	36%
In-your-care time, overlapping with leisure	38%	38%	40%
In-your-care time, overlapping with leisure and unpaid housework and shopping	76%	75%	76%

One might argue that the marginal cost of supervision is small if the respondent plans to perform housework. Still, the causality often runs the other way, i.e., a parent who must be at home to supervise children is more likely to perform housework or shop because engaging in other productive activities is impractical.[10] Given the existing structure of wage rates that we use in our valuation, the implication is that combined housework–supervisory care time would be valued more highly than active childcare and adult care. It is difficult to say whether this form of multitasking delivers the same quality as solo tasking or whether it lowers the quality and should therefore be somewhat discounted. From a consumption-value perspective, it may be argued that the combined hour should be valued, if we assume no significant

[10] This has changed since 2019 in the wake of increased adoption of technologies that facilitate telecommuting.

reduction in quality, at an even higher rate than we propose. Like other decisions regarding replacement wage rates, this one is based more on intuition than evidence and deserves further consideration.

Finally, adult care often entails responsibilities—including medical supervision of frail elderly or others experiencing long-term disabilities—that require commitment and management skills. Like childcare workers, elder care workers are generally considered underpaid in the US, especially when employed in home- and community-based care where labor standards are poorly enforced. As with childcare, we have good reason to believe that unpaid care provided by family members entails person-specific skills, one-on-one attention, and social-emotional benefits, all of which should be considered in its valuation. For these reasons, we chose an occupational category that, like that chosen for specialist valuation of childcare, is a step up from the most poorly paid, high-turnover occupation of "Home health and personal care aides" (OEWS code 31-1131). The best candidate may be "Nursing assistants" (OEWS code 31-1120). These workers "perform duties such as monitoring health status, feeding, bathing, dressing, grooming, toileting, or ambulation of patients … medication administration and other health-related tasks."[11] National estimates for 2021 from OEWS data show that the average hourly wage was $14.07 and $15.99, respectively, for home health aides and nursing assistants. We chose the latter for our valuation of adult care.

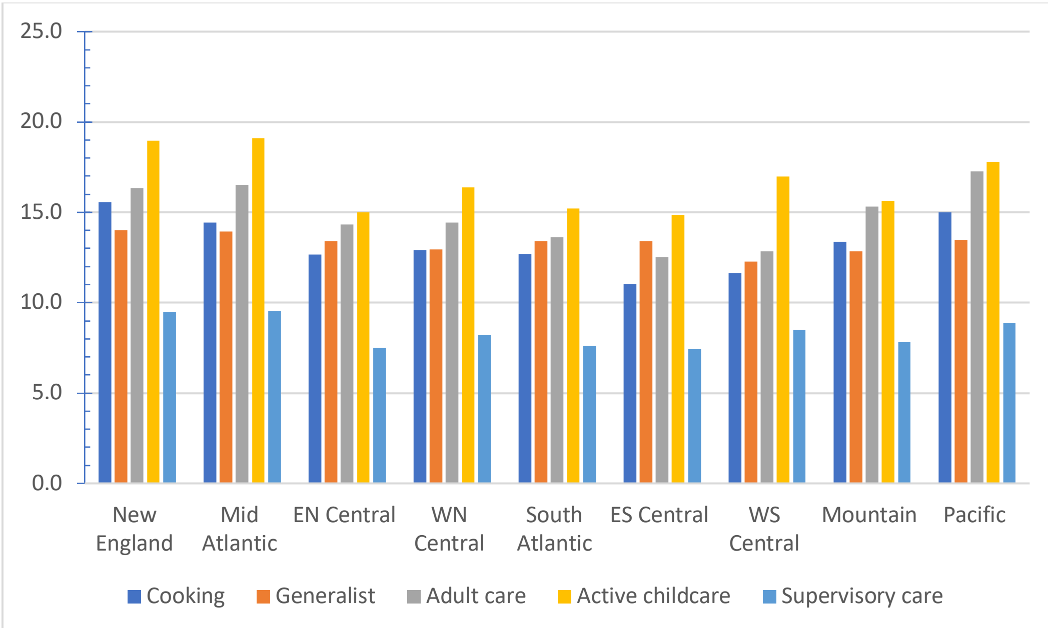
We also include care received from people outside the household in our definition of household production (Table 2-1). We apply the same rate for childcare for active childcare rendered by a household member. Similarly, we value care for adults given by people outside the household at the same rate as members of the home.

We constructed an alternative set of estimates to assess the sensitivity of our preferred definitions. They exclude supervisory care and care received from people outside the household. Further, the valuation of the remaining activities of household production shown in Table 2-1 was performed using the generalist wage. Because the scope of household production and valuation method in these alternative estimates are similar to the BEA methodology, we refer to them as the "BEA method" in our comparisons (Chapter 5).

As with many occupations, there is considerable geographical variation in the wages of the occupational groups we discussed (Figure 2-1). Therefore, we recommend using at least state-level wages in the CE internal data. In the public-use data, the Census division is the lowest level of geography available for all records. Consequently, our valuation exercise used the average wage for each occupation, differentiated by division.

[11] This description is from Occupational Employment and Wages, May 2021. Nursing assistant definition is available at: https://www.bls.gov/oes/current/oes311131.htm

Figure 2-1 Hourly wage rates used in valuing household production by category (2019 dollars)



As expected, wages for all categories are generally higher in the New England, Mid-Atlantic, and Pacific divisions, compared to other areas. In all divisions, the average wage for active childcare is the highest, followed by that for adult care. We found the gap between the two wage rates to be rather small in three of the nine divisions: East North Central, Mountain, and Pacific. Two wage categories that correspond to housework—cooking and generalist—do not show a clear ranking across divisions. In the New England and Pacific divisions, cooking has a clear lead over generalist, while in East South Central, the reverse holds. Across the other six divisions, a notable difference between the two is absent. An anomaly is presented by the East South division, where the adult care wage is below that of the generalist. The lowest wage in all divisions is that assigned to supervisory care, which is half of that for active childcare. Hence, the variation across regions is the same for both categories of childcare.

Notwithstanding the regional diversity, the national pattern of our chosen wage structure ranks active care work in households higher than household chores, and supervisory care ranks the lowest. Examining the range of wages across categories gives us a fuller picture regarding valuation. The wage for cooking falls between \$11–\$16 across regions, while the generalist wage has the tightest band with a range of \$12–\$14. The ranges for the two care categories are \$13–\$17 and \$15–\$19 for adult care and active childcare, respectively. Given the closeness of the wage band for active childcare to that of other categories, and our method of valuing supervisory childcare at half of that for active childcare, we end up valuing an hour of any class of household production (cooking, other housework, and adult care) that overlaps with supervisory care at a higher rate than an hour of active childcare. Since active childcare has the highest wage, the combined effective wage of supervisory care that overlaps with household production will also be higher than for any other category of household production.

2.4 Summary

We have outlined the list of nonmarket activities that need to be brought into the picture in a broad measure of consumption. Our scope of activities differs from the standard definition used in most previous work (e.g., by the BEA), in that we account for supervisory care or in-your-care time with children and care received by household members from people outside the household. These extensions are warranted because they contribute to households' potential consumption. We also specified a valuation procedure to convert household production hours into monetary units. In contrast to the current BEA practice of using the oldest and most widely deployed generalist wage method, we favor the specialist wage approach, which, in fact, was used by earlier research on incorporating household production in US national accounts (Landefeld, Fraumeni, and Vojtech 2009).[12] Our specialist wage structure assigns a greater hourly value to active care for dependents than to housework.

[12] However, still earlier work by Landefeld and McCulla (2000) used the generalist wage.

3 Imputing Home Production by Members of the Consumer Unit

We begin this chapter with an overview of the methods previously used in the literature on imputation. Our focus is on studies attempting to impute the time spent on household production, especially on efforts to combine time use and consumption expenditures. The subsequent section outlines the methods of imputation we use in the project in a more or less formal manner. We do not aim to provide all the technical details since our techniques are relatively well known, and excellent treatments of the various methods are readily available elsewhere. Instead, our exposition is biased toward those who may not be immersed in the literature on imputation. The following section is the longest. We present details on the implementation of imputation methods using the datasets for this project. Alternative methods are also assessed for the quality of imputations that they generate, in other words, how "closely" they can mimic the features of the observed, i.e., time-use survey, data. We put the term "closely" in quotes because defining closeness involves ingredients of judgment and emphasis rather than formulaic definitions, as we shall see below. A validation exercise using data from the Panel Study of Income Dynamics is discussed in a companion appendix to the chapter (Appendix A). As discussed there, all imputation methods rely on an assumption known as the conditional independence assumption (CIA). A direct test of this assumption is generally impossible; our validation study offers the next-best alternative, given the currently available data.

3.1 Background

Previous research has combined microdata from time-use surveys and income or expenditure surveys for various research goals. Perhaps the earliest instance for the United States is that of Fuchs (1986), who attempted to estimate total hours of work, leisure, and income for men and women separately using the 1975–76 time-use survey (TUS)[13] and the public-use samples from the decennial censuses of 1960, 1970, and 1980. The method used by Fuchs for his 1980 estimates consisted of two steps: (1) estimating an ordinary least square regression (OLS) of hours of nonmarket work or household production on a set of explanatory variables in the TUS, and then (2) fitting the estimated model to the Census data. Fuchs obtained the OLS estimates of the core subgroup of interest to him—the working-age population (ages 25 to 64)—separately for men and women. The explanatory variables were the dummies for ages 45–64, Black, married, full-time employment, part-time employment, and the presence of a child under age 5 in the household. For those ages 18–24 and 65 and over, the OLS estimates were obtained with age,

[13] This survey was conducted by the Institute for Social Research at the University of Michigan. See Juster et al. (1979) for documentation.

dummies for sex, marital status, and employment status as independent variables (Fuchs 1986: S270–S271).[14]

A recent application of the same method is worthy of mention in our context because it involved imputation of time use in the CE. The study sought to estimate the joint distribution of family-level per capita leisure and consumption using time-use and CE for four years between 1975 and 2016 (Han, Meyer, and Sullivan 2020). Family-level leisure and consumption are defined as weekly leisure per adult and yearly amount per equivalent adult, respectively. In addition to the 1975–76 TUS mentioned above, the authors use the 1985 unofficial TUS as well as the 2003 and 2016 rounds of the ATUS. Leisure time was imputed for individuals in the CE in the corresponding years (except that the 1975–76 TUS is paired with the 1972–73 CE). Family-level leisure is computed as the average for adults in the household. OLS estimates used in the imputation are the coefficients for the independent variables by year obtained by pooling the various time-use surveys. The independent variables used in the analysis included individual characteristics (dummies for age group, sex, educational attainment, and marital status; and usual weekly hours of employment) and family characteristics (number of children under five, number of children under 18, number of adults). An important difference from Fuchs (1986) is that Han et al. (2020) attempt to control for within-year variation ("transitory variation") in leisure by including interview-month dummies and interactions of the year with interview day of the week as regressors.[15]

A modification of the "fit and predict" method of imputing household production was introduced by Jenkins and O'Leary (1996) in their study of the distribution of extended family income (standard money income plus the value of household production) in the UK (see Frazis and Stewart [2011] for the US). The modification consisted of adding an individual-specific random number drawn from a normal distribution with mean zero and variance set equal to the estimated variance of the error term in the model fitted in the TUS, viz., the Time Budget Extension of the 1987 ESRC Social Change and Economic Life

[14] Schor (1992) employed the same method for imputing hours of household production in the March CPS supplements of 1970, 1974, 1980, and 1988. She used the TUS used by Fuchs (1986) and the 1981 wave of the same TUS. OLS estimates were obtained for women and men separately. The regressors included dummies for age group, marital status, housewife, head of household or spouse of the head of household, Black, and other race, as well as family income, hours of employment, square root of the number of children under 3 years, and square root of the number of children above 3 years (Schor 1992, 171–74).

[15] Apps and Rees (1996), while deploying the same method of imputation in the Australian Bureau of Statistics (ABS) 1985/86 Income Distribution Survey with the OLS coefficients estimated from the 1987 ABS Time Use Pilot Survey, expanded the list of regressors in a notable fashion. In addition to household characteristics and own characteristics of the respondent, they incorporated characteristics of *other* household members, specifically, the married or cohabitating partner, such as their hours of employment and dummies for their occupational group (see Apps and Rees 1996, 217–18). The inclusion was prompted by the authors' goal to compare the parameter estimates from two variants of the collective labor supply models: the "transfer model" which neglects the intrahousehold division of housework (by treating housework and leisure identically) and the "exchange model" that explicitly incorporates the gender division of housework.

(SCEL) survey. OLS estimates were obtained from the TUS separately for four groups of adults differentiated by sex and marital status. The regressors comprised respondents' age group, employment status, and household composition. These estimates were then used to impute household production hours in the 1986 Family Expenditure Survey. However, the authors note that the modification of the method did not alter their most important conclusion that adding household production had an equalizing effect on income distribution (Jenkins and O'Leary 1996, 407n15). They also point out that the crucial assumption required to justify the imputation—the conditional independence assumption (CIA)—is hard to test in practice. As a substitute, they calculated the correlations between the total time spent by the household on domestic work and the poor-quality data on household money income for each value of the regressor variables used in the TUS regressions. Most of the correlations were not statistically different from zero, and they concluded that the assumption "may be a satisfactory approximation in the current context" (Jenkins and O'Leary 1996, 408).

Kolodinsky and Goldstein (2011) represent a nonparametric approach to examining the relationship between obesity, time use, and food expenditure patterns in the US. Since no one survey contains all three variables, the authors turn to imputation to overcome the information gap. Specifically, they pool the data from the Eating and Health Module of the ATUS for 2006 and 2007. This module collects information on the height and weight of individuals. From this information, the authors compute the body mass index (BMI), which they use to measure obesity. Kolodinsky and Goldstein (2011) also utilize the demographic information on the respondents that is directly linked to their CPS record and available as a part of the ATUS. Their analysis focused on single-female household heads ages 31–50 and without disabilities. They use a nonparametric conditional-mean imputation approach to obtain the food expenditure pattern for the selected households. Specifically, they assign average values of food expenditures by detailed categories from the CE 2007 to the ATUS, using cells that are defined by gender, income, and family composition. As part of their statistical analysis, the authors model the likelihood of being obese as a function of demographic characteristics, time-use activities (related to activities or eating), and specific categories of food expenditures. The effects of (imputed) expenditure variables are identified in their statistical model only because imputed data was transferred nonparametrically (cell means), and no corrections were applied to standard errors in their analysis.

Recently, researchers in the economics of transportation have also realized the importance of data that contains information on time spent on various activities and expenditures on associated categories of commodities that may complement, substitute, or serve as inputs for such activities. For example, Konduri et al. (2011) developed an a-theoretical statistical model that estimates the willingness to pay for leisure activities and a reduced-form linear expenditure system (the so-called Stone-Geary model) to ascertain the (opportunity-cost) value of leisure. Their empirical analysis uses the sub-samples of households with only one person who is also employed from the 2008 rounds of the ATUS and CE (Interview sample). The authors argue that a week is the appropriate period of analysis because expenditures of time and money follow weekly cycles (Konduri et al. 2011, 56). However, the ATUS collected the time diary for a single day from each individual in the sample. To overcome this limitation, the authors impute entire time diaries for each respondent for days other than their interview day. The imputation method is to take random draws with replacements from the sample for all days of the week (except the day of the interview), using a pool of donors based on matching demographics (such as gender, age, employment status, race, college status, family income, and employment category). Similarly, they also imputed expenditures to the ATUS respondents by conducting random draws with replacements from the CE sample, with the only difference being that the employment category was dropped from the list of matching characteristics. Weekly expenditures in the CE sample were created by converting quarterly expenditures via "naïve scaling." While the authors describe their method of matching as simplistic, they do not adjust their statistical analysis to account for the imputation and consider the imputed variables on the same footing as the original variables in the ATUS.

Another example of combining expenditures and time use in the literature on transportation economics is Hawkins and Nurul Habib (2022). They attempt to create a synthetic data file by combining household-level expenditure data with individual time-use patterns, which should be statistically valid for urban simulation models. The authors argue that the main limitation of standard data fusion/statistical matching approaches is that they obfuscate spatial information, linking observations from different locations. The authors apply their proposed framework for the Greater Toronto Area in Canada, fusing data from the Canadian Census of Population (CCP 2016) for general demographics, the Survey of Household Spending (SHS for years 2012–16) for expenditures on goods and services, and the General Social Survey: Time Use Survey (TUS 2010 and 2015) for time spent on activities. In contrast with other methods, which use one of the surveys as baseline/recipient, transferring the required data from other surveys after matching, their approach uses the different surveys to identify the joint distribution of all variables of interest and create a fully synthetic dataset based on the information from the reference surveys. To do this, the authors utilize generative modeling algorithms borrowing from machine learning and Bayesian literature, specifically the algorithm proposed by Ye at al. (2017). A peculiarity of this approach is that there is, in effect, no "recipient" data file. They discuss the potential limitations and care required to consolidate the different time frames for data collection across surveys. Hawkins and Nurul Habib (2022) also acknowledge that not having access to time use for every day of the week, their methodology may ignore the correlation across substitutable day-to-day activities. The authors compare the distributions of the analysis variables in the synthetic dataset using—for time use—data from the Transportation Tomorrow Survey (TTS-2016) and the CCP for sociodemographic patterns.

Kum and Masterson (2010) provide a review of statistical matching, describing different methodological choices that would link statistically similar individuals across surveys, thus allowing for the creation of synthetic datasets with information from both surveys. They discuss the matching algorithm used to construct the Levy Institute Measure of Economic Well-Being. The authors match records from the Survey of Consumer Finances (donor file) with records from the March supplement of the CPS using a rank of predicted propensity scores in a constrained matching procedure with strict strata variables to improve the quality of imputation. In contrast to other methods described above, the procedure adjusts survey weights to achieve almost perfect alignment and transfer the full wealth distribution from the SCF to the CPS. They also advocate for the use of sample segmentation or stratification. This stratification helps improve the match quality by restricting the potential donors and recipients across both files. Distinct from other methods, the segmentations are dynamic, so the definition of matching groups is expanded after all possible observations within the more detailed segment definition have been utilized. The authors' evidence suggests that statistical matching can transfer overall distributions almost perfectly to the donor survey, with minor losses when using some of the same variables used for the sample segmentation. They warn, however, of potential problems associated with statistical matching when the survey structures between donor and recipient are too dissimilar, as it may produce attenuation of inequality in the donor dataset.

As described below, our preferred method of imputation for the current project is a variant of the Kum and Masterson (2010) procedure, linking information from the time-use survey (donor) to the Consumer Expenditure Survey (recipient). The modifications include splitting survey weights to integrate the two files fully. We also deploy principal-component analysis and cluster analysis to pre-identify similar groups in the two files and improve the imputation quality.

3.2 Imputation methods

As described in Chapter 2, our definition of household production encompasses unpaid services rendered by household members and nonhousehold people for household members. Regarding time expenditures, services by own-household members far outweigh those by nonhousehold members. In our judgment, the data available for imputing the latter is also of lower quality—a point we elaborate on in the next chapter—than that available for the former, namely, the ATUS. Consequently, we produced imputations with three methods using the ATUS to allow choices for the data users depending on their research goals.

Only one approach—the most direct and widely used—was deployed for imputing the services provided by nonhousehold members (see Table 3-1).

Table 3-1 Imputation methods used by the component of household production

Component of household production	Imputation method(s)
Household production for members of own household	Regression prediction (RP), Multiple imputation (MI), and Statistical matching (SM)
Nonmarket care of older adults received from nonhousehold members	Regression prediction with noise (RP+)
Nonmarket childcare received from nonhousehold members	Regression prediction with noise (RP+)

The methods we used belong to one of the two families: imputation by prediction or imputation by matching. We outlined a few examples of both approaches earlier in this chapter. We begin with a description of the prediction method and then turn to statistical matching.

Our first variant of the prediction approach uses a univariate Tobit regression model where we assume that hours (x_{ij}) spent by individual i on activity j (e.g., time spent on childcare) can be modeled as follows:

$$x_{ij} = \max\left(0, z_i \beta_j + e_{ij}\right), \quad j = 1,2, \dots, k.$$

where $e_{ij} \sim N\left(0, \sigma_j^2\right)$ and z_i is the set of characteristics (e.g., age of the respondent) and β_j is the corresponding set of unknown regression coefficients. This approach is usually appropriate given the substantial number of people (particularly among men) who spend no time on household production during a week, which translates to a dependent variable containing a sizeable number of zeros. In the next step, we assign imputed values to each individual using the overall expected mean value of hours, conditional on the control variables:

$$E\left(x_{ij} \middle| Z = z_i\right) = E\left(x_{ij} \middle| Z = z_i, x_{ij} > 0\right) P\left(x_{ij} > 0 \middle| Z = z_i\right), j = 1,2, \dots, k$$

The second variant creates imputed values for hours based on the theoretical data-generating process:

$$\hat{x}_{ij} = \max\left(0, z_i \widehat{\beta_j} + \hat{e}_{ij}\right), j = 1,2, \dots, k$$

Where $\hat{e}_{ij}$ is a random draw from a normal distribution with mean zero and standard deviation $\widehat{\sigma_j}$. The hope is that adding the errors, $\hat{e}_{ij}$, provides a stochastic component that brings the distribution of the imputed variables closer to the true but unobserved distribution of household production time. Following Crossley et al. (2020), we refer to the first variant as "RP" (regression-prediction) and the second variant as "RP+." As indicated in Table 3-1 above, the RP method was used for the various categories of household production services provided for members of their households. At the same time, we implemented the RP+ method for, respectively, nonmarket care services for children and older adults received from people outside the home.[16]

The third variant we implemented is multiple imputation (MI) which is generally regarded as a better choice for conducting multivariate analysis and statistical inference. It has been argued that, because the MI procedure provides several imputed values of the same variable for each record (e.g., five different values are generated of the time spent on cooking for each person), imputation errors can be accounted for and confidence intervals calculated for conditional means or regression coefficients (see, e.g., van Buuren 2007, 219). Given current software capabilities, we employ two variants to implement MI.

We used the MI procedure in SAS with a fully conditional specification (FCS) and regression method as one variant. This approach, however, assumes the outcome can be modeled as a linear function of characteristics. Given the nature of our data, a method that explicitly allowed for censoring (e.g., Tobit) would have been preferable; unfortunately, SAS currently does not provide such an option. A linear regression model may not produce usable imputations because it may produce negative values for the imputed data. In addition, applying a censoring or correction rule to the imputed data (which would zero out any negative imputations) may produce severely biased results with a conditional mean larger than the observed data mean.

To assess the sensitivity of the imputation to censoring, we carried out MI using Stata, which allows an interval regression model for the variables to be imputed (see Royston 2007 for a discussion of the method). Interval regression is a generalization of the Tobit model, which allows for the arbitrary structure of censoring and, in principle, follows a logic similar to the RP+ described earlier. In our implementation, we assume that positive values for all time-use categories are fully observed and uncensored. Furthermore, cases with zero hours are supposed to be interval-censored, with zero as the upper bound and $-\infty$ (unbounded) as the lower bound.

[16] It should be noted, however, that the RP+ methodology applied for care received from outside the household assumes distribution functions other than the censored normal distribution.

In Appendix B, we conduct a preliminary comparison of the results from the two procedures. We expected that the Stata procedure would produce much better results because it explicitly accounts for the censoring. However, the results are somewhat more ambiguous. Despite this, we think that a procedure similar to the Stata procedure should be used because it is conceptually more appropriate.

As discussed, e.g., in Enders (2022) and Newman (2014), MI overcomes some of the limitations of the RP and RP+ approaches. In particular, it permits unbiased estimates of coefficients and standard errors under the missing-at-random (MAR) assumption.[17] In addition, because the method imputes multiple values for the structural component $z\beta$ and the unobserved component ε, the imputed data is likely to reproduce moments of the observed distribution of the variables in the donor file. Of course, the validity of the MI procedure, just as in all prediction methods, depends critically on the assumption that the parametric model accurately describes the data-generating process. In contrast, the statistical matching (SM) method we prefer is nonparametric and does not impose any distributional assumptions for the imputation procedure.

The statistical framework of our procedure has been described in detail elsewhere (Kum and Masterson 2010). In brief, we employ constrained matching based on propensity scores and stratification. The constraint is that matching is done without replacement (i.e., all the records in the donor file are used) to equalize the weighted number of observations in the donor and recipient files. To ensure equality, we inflate weights in the donor file by the ratio of the sum of weights in the recipient file to the sum of weights in the donor file. This transformation allows all donor records to be matched to recipient records by splitting their weights as needed (see below for weight splitting). Once this transformation and other steps in data harmonization are carried out (e.g., making sure that the age variable in both files has the same range), the data files are stacked to form a single file. To estimate the propensity scores, we define a binary (outcome) variable, T, that takes a value of 1 if the record is in the recipient file and 0 if in the donor file. We then partition the data into subgroups (e.g., women with children) and estimate, for each observation and subgroup, propensity scores based on a logit model:

$$P(T = 1|Z = z_i) = \Lambda(z_i\beta) = pscore_i,$$

where Z is the set of covariates and β is the associated set of parameters. The propensity scores are obtained by modeling donor/recipient membership at the highest level of aggregation (i.e., for the whole stacked file) and for the main subgroups (e.g., subgroups formed by gender and presence of children). To further improve matching quality, smaller subgroups are created and identified using cluster analysis and

[17] According to Enders (2010), data Y are said to be missing at random (MAR) when the probability of being missing depends on other variables, and not on itself.

principal component analysis (PCA) (as described in the later section on the matching process). Still, they rely on propensity scores estimated at the highest level of aggregation. We employ PCA as a data reduction technique to obtain a few principal components representing a more extensive set of observed characteristics. These principal components are then utilized to identify different groups of observationally similar individuals by applying a modified partitioning cluster algorithm. In turn, these subgroups are deployed as sub-strata in the statistical matching procedure.

We match observations based on rank similarity within subgroups. Thus, we sort the observations in each file by subgroup based on the estimated propensity scores. A record with a larger weight in the donor (or recipient) file will be split up or duplicated and matched with multiple records in the recipient (or donor) file until all of its weight is used. In practice, however, subgroups will not be of the same weighted size, and the weight-splitting will not completely match the records in both files using the most detailed subgroups. Thus, additional steps (i.e., using higher levels of aggregation for subgroups) will be required to accomplish this goal. Because multiple "donors" could be linked to a single "recipient" observation, and vice versa, we choose only one of all the possible candidates as the effective donor. We select the donor observation matched in the earliest round, representing the largest "share" of the weighted recipient observation.[18]

3.3 Imputation results

3.3.1 Household production for members of own household

3.3.1.1 Introduction

The main category of household production we impute is time spent on household production by members of the household for their household. Our source data or donor file is the 2019 American Time Use Survey (ATUS), which collected 24-hour time diaries from a sample of individuals aged 15 years or older. We imputed hours for each individual 15 years or older in the Consumer Expenditure Survey (CE). Imputations were carried out for the 2019 CE Interview and CE Diary samples, using each quarter separately.[19] Separate imputation is warranted because the Interview and Diary samples are independent,

[18] For example, if recipient r1 is matched with donors d1 and d2 in rounds 3 and 5, donor d1 is selected as the effective donor. If recipient r1 is matched with donors d1 and d2 in the same round, the donor with the largest split weight is chosen as the effective donor.

[19] For the Interview sample, we separately matched memi191x, memi192, memi193, memi194, and memi201 data files. Separate matches were also conducted for the Diary sample files, memd191, memd192, memd193 and memd194. We used information from the corresponding fmli files to construct household-level variables (such as family income) to be used in the matching. For each of the CE recipient files, the full-year ATUS 2019 was the donor file.

and each quarter is considered independent. It is also desirable for several analytical purposes of the matched data because users may want to consider only one of the samples. For example, relatively more frequent expenditures, such as food expenditures, are captured better in the Diary sample. Users may want to examine the correlation between those expenditures and time spent on housework or shopping.

We explored two options for combining the information from the two surveys. One procedure involves treating all the records in the ATUS simultaneously as donor records, while conducting separate imputations with the Interview and Diary samples. A potential drawback of this procedure may be seen by considering that the ultimate objective of the exercise is to develop household-level estimates of the value of household production. The procedure has no explicit way of preventing, e.g., in the context of the statistical matching, all or a majority of individuals in a CE consumer unit from being assigned only weekend or weekday diaries. Such a result would mean we do not have a reliable picture of household production during the average day in the week of the household. Further, the mix of weekday and weekend records in consumer units can unpredictably vary across subgroups of households, thus making intergroup comparisons problematic.

The alternative procedure entails treating weekday and weekend diaries as separate donor files. Each record in each CE sample is then imputed with weekday hours and weekend hours separately from the respective donor files. We can then calculate the amount of time spent on household production during the average day of the week by the individual in a given CE sample as a weighted average of the time imputed from the weekday and weekend matches, respectively. The averages thus calculated for individuals in the household can then be added up to obtain the household-level total of time spent on household production, avoiding the potential bias from treating the weekday and weekend diaries as a single donor file. We consider this second procedure to be the most appropriate, as it would better capture the potential heterogeneity in time allocation between weekdays and weekends. [20]

A crucial assumption underpinning all imputation methods is that the donor and recipient files represent the same underlying population. Therefore, we compared the demographic picture conveyed by the ATUS and CE samples to assess their mutual correspondence. Given the differences in sample design, survey implementation, and sampling variance, we do not expect the picture to be identical, yet, a close

[20] In the ideal scenario, one would prefer using each day of the week as a separate matching task, so that we could better reconstruct the time-use heterogeneity across a typical week, month, or year. However, a potential drawback of treating the weekday and weekend diaries as separate donor files is that the size of the donor pool is cut in half because the time diaries in the ATUS are split equally between weekdays and weekends. The smaller donor pool can affect the quality of the imputation, because each ATUS subsample may be less representative of weekend or weekday time-use activities. Using weekend and weekday time–use data separately is a good compromise, given the time-use heterogeneity, although we do not recommend using a lower-level subsampling (for example, using quarter or month).

resemblance is desirable and necessary for effective imputation. Table 3-2 provides the total number of observations across each survey and the total number of weighted observations.

Table 3-2 Weighted vs. actual observations across surveys (individuals aged 15 years and older)

Sample	Weighted Obs.	N
CE Interview		
memi191x	262,286,293	10,861
memi192	261,946,747	10,630
memi193	263,038,925	10,357
memi194	263,864,691	10,158
memi195	263,070,970	10,124
CE Diary		
memd191	262,018,321	5,156
memd192	262,356,969	5,353
memd193	263,622,588	5,380
memd194	263,390,712	4,952
ATUS		
ATUS Weekday	263,653,004	4,642
ATUS Weekend	262,635,719	4,793

Roughly speaking, all samples represent approximately the same number of individuals (263 million). The actual sample sizes are different, and the ATUS sample is considerably smaller. In the context of statistical matching, the differences in size imply that each ATUS (weekend or weekday sample) record may be used as a donor record, on average, for up to two observations in the CE Interview sample. Consequently, the variances of the imputed conditional distribution we may observe in the CE data may be a slight understatement of the "true" but unknown conditional distribution.[21]

As described below, gender, presence of children in the household, and employment status are crucial in our imputation exercise. They are also widely recognized in the literature as driving factors behind variations among individuals in the time spent on household production. A simple comparison of the gender-wise distribution of employment status and children's presence in the household across all samples is shown in Table 3-3. We report average characteristics across all quarters for the CE rather than consider each quarter's statistics separately. The statistics suggest a reasonable balance across all samples regarding this particular classification of individuals. Additional tables comparing the distribution across a more extensive set of variables used in the analysis can be found in the online data appendix.

[21] In this case, conditional distribution refers to the distribution in time use we would expect to see among households with exactly the same characteristics.

Table 3-3 Composition of the sample by sex, employment status, and presence of children in the household (percent)

	CE Interview	CE Diary	ATUS Weekday	ATUS Weekend
Men				
Not employed; no children	20.58	20.46	20.61	22.11
Not employed; with children	8.39	9.08	6.83	7.80
Employed; no children	42.09	42.51	44.77	43.93
Employed; with children	28.94	27.95	27.79	26.16
All	100.00	100.00	100.00	100.00
Women				
Not employed; no children	25.61	26.83	27.68	25.88
Not employed; with children	14.03	12.97	11.64	12.59
Employed; no children	34.76	34.50	35.65	36.95
Employed; with children	25.59	25.70	25.03	24.58
All	100.00	100.00	100.00	100.00

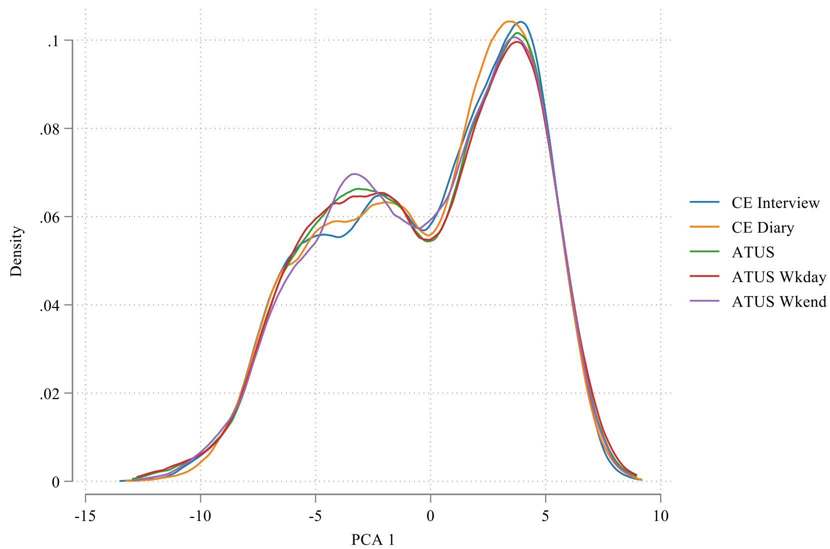
Note: CE Interview and CE Diary correspond to pooled data across all available quarters.

To summarize the findings from the detailed tables provided in the online appendix: there is a close correspondence of the critical characteristics across all surveys, with a few cases where the correspondence is less than desirable. We describe a few of these cases:

- In the CE Interview and Diary samples, there are about 3 percentage points fewer individuals with a GED and 2 percentage points fewer individuals with an advanced degree (Master, Ph.D., or similar) compared to the ATUS, but a greater presence of individuals with only some college education (3–4 percentage points higher)
- A slightly larger proportion of individuals in households with children in the CE than in the ATUS.
- A higher share of homeowners (4.5 percentage points more) in the ATUS than the CE.
- There are some differences in the distribution of individuals across annual income levels, especially for those in the $100K–150K bracket.

While these differences may seem notable individually, the overall differences in sample composition are not as pronounced. For instance, Figure 3-1 plots the kernel density of the first principal component built on all selected characteristics, which is used later for matching. Since a principal component is a linear combination of characteristics that explain most of the variation of the data, it can be used as a simple approach to compare the balance across two datasets. This use of principal component analysis is similar to applying propensity scores for matching and balancing assessments in the policy evaluation literature. Although some differences in the densities exist, they follow each other closely, suggesting good balance across all surveys.

Figure 3-1 Density of the first principal component



Note: For the CE Interview, and CE Diary, the plot corresponds to all quarters' pooled information

3.3.1.2 Imputation by matching

3.3.1.2.1 Variables and cell construction

We describe the variables used to implement the matching algorithm and propensity score estimation in Table 3-4. Propensity score estimation was conducted by interacting all variables with gender and the presence of children. We can classify the variables used in the process into three groups: principal strata, secondary strata, and others. A principal strata variable is kept in all rounds of matching. In our application, all statistical twins in the match are defined within the cells "men with children," "men without children," "women with children," and "women without children." We retained secondary strata variables in the first 12 rounds of matches, which, as shown below, encompass most observations. We chose these strata variables because the differences in household production among these subgroups are known to be sizeable. Finally, variables classified as "Other" are used more flexibly in the matching process to define subgroups.

Table 3-4 Variables used in matching and propensity score estimation

Variable	Description
Principal Strata	
Gender	Dummy for male/female
Co-residence with children	Dummy for any child 0–17 in the household
Secondary Strata	
Employment status	Dummy for whether a person is currently employed or not
Number of children (0,1,2,3+)	Dummies for the number of children in the household
Other Variables	
Own Characteristics	Interviewed person
Age	Age in years
Education level	Level of education
Self-employment status	Dummy for self-employment
Race	White, Black, Hispanic, and other
Householder 1's characteristics	Husband in a couple-household, older male or female in a household with a single reference person, and male or female in a same-sex household
Gender	Dummy for male/female
Age	Age in years
Education level	Level of education
Employment status	Dummy for whether a person is currently employed or not
Self-employment status	Dummy for self-employment
Race or ethnicity	White, Black, Hispanic, and other
Householder 2's characteristics	Wife in a couple-household, younger male or female in a same-sex household, in a household with a single reference person
Gender	Dummy for male/female
Age	Age in years
Education level	Level of education
Employment status	Dummy for whether a person is currently employed or not
Self-employment status	Dummy for self-employment
Race or ethnicity	White, Black, Hispanic, and other
Household Characteristics	
Couple/single household	Type of household (Single headed or couple head)
House tenure	Owner or renter
# Children 0–5	Number of children in the household between 0 and 5 years of age
# Children 6–12	Number of children in the household between 6 and 12 years of age
# Children 13–17	Number of children in the household between 13 and 17 years of age
# Male Adults 18–64	Number of male adults in the household
# Female Adults 18–64	Number of female adults in the household
# Older Adults 65+	Number of older (65+ years) adults in the household
# Employed Adults 18–74	Number of employed adults in the household
Family income group	Family pre-tax income group (based on ATUS classification)

In addition to the primary strata variables (gender and co-residence with children) and secondary strata (employment status and number of children), we employ cluster analysis to identify detailed subgroups to perform the statistical matching. The analysis is done in two steps. In the first step, within each primary strata subgroup, we use all the variables used to estimate the propensity scores and apply principal component analysis (PCA) to reduce the dimensionality of the explanatory variables.[22] We estimate eight principal components from the PCA, all of which have an eigenvalue above 1.

Once the principal components have been obtained, they are used to implement a cluster analysis, where individuals are assigned to a predefined number of groups to minimize the intra-group dissimilarities and inter-group similarities. We aim to find homogeneous subgroups based on observed characteristics. To implement this, we use a $K\text{-}means$ partition cluster algorithm because of the computational advantages over hierarchical clustering algorithms.

Because partition algorithms are known to identify only locally optimal clusters, we use an extended algorithm, aiming to find globally optimal clusters. For a given number of predefined subgroups, we identify 20 cluster candidates and estimate the Calinski-Harabasz $pseudo\text{-}F$ index (Calinski and Harabasz 1974) for each cluster. The cluster with the highest index is selected as the optimal cluster. We perform the cluster selection independently within each primary strata group and identify optimal clusters constituting 10, 20, …, 100 homogenous subgroups. For the most detailed strata, we combine information on primary strata, the number of children and employment status, and 100 groups of clusters for their construction. The result is 1,502 subgroups compared to 4 subgroups when only the primary strata variables are considered.

3.3.1.2.2 Process

Given the sample sizes of the donor and recipient surveys, we were concerned about the number of observations in the cells formed based on the procedure described above. Therefore, we modified the procedure to avoid matching within "small" cells, i.e., cells containing only one or two observations from a given sample. Specifically, we constrain the match such that cells containing less than 0.1 percent of total observations in the data are ignored. In other words, we required at least 5 observations from the ATUS weekday and weekend samples, 10 from the Interview sample, and six from the Diary sample within a cell to use a particular cell for matching. The impact of the modification can be seen by considering the summary of the matching between the Interview sample and the weekday and weekend diaries, respectively (Table 3-5 below). Specifically, about 74 (Interview) and 70 (Diary) percent of the

[22] We do not use secondary strata for the identification of subgroups via cluster analysis because the cells would have been too small to implement the partition-cluster algorithm.

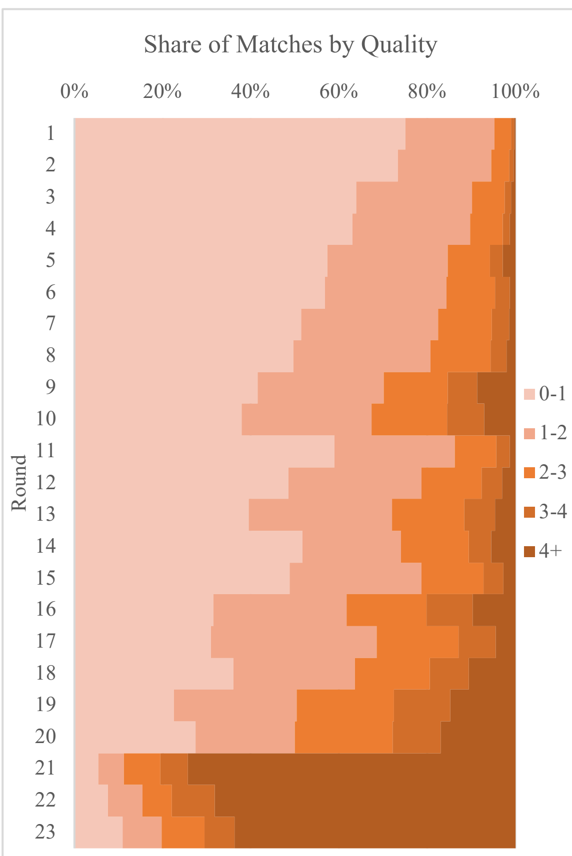
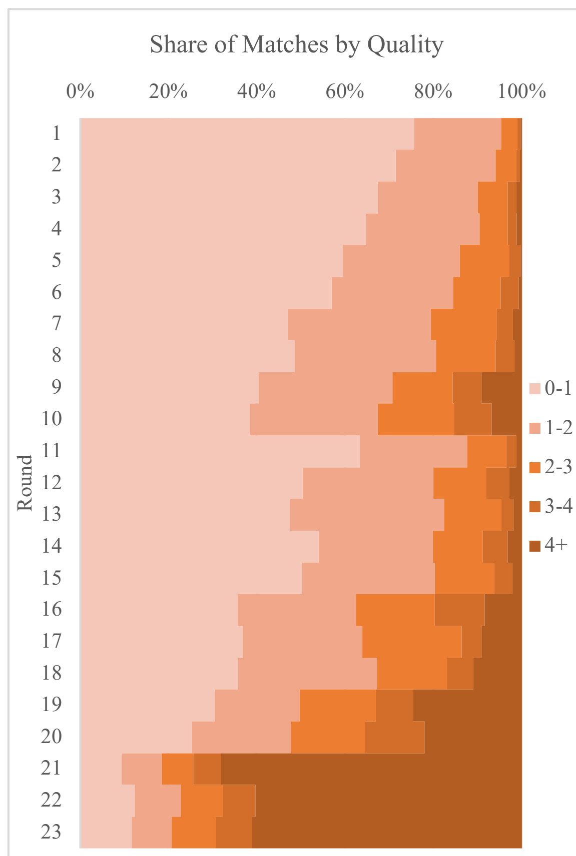
data is matched in the first round when small cells are used, compared to 47 percent when small cells are not allowed. While this may suggest that imposing the cell-size constraints affects the match quality negatively, one should also consider that using those micro-cells may be equivalent to overfitting, which can negatively affect the overall match quality. However, as shown in Figure 2, there is little loss in terms of matching quality when imposing the cell-size restriction.

Table 3-5 Share of the total number of observations matched in each round (percent) in the matching of the CE Interview sample with ATUS

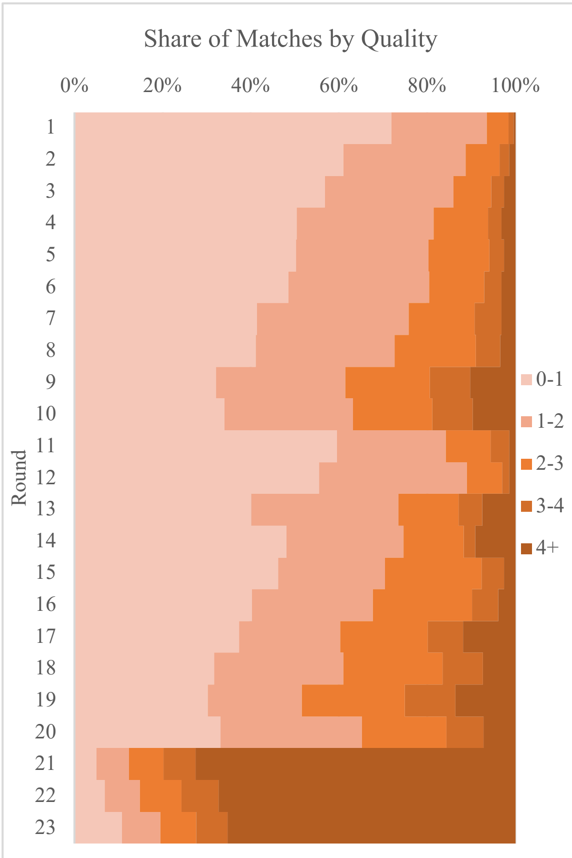
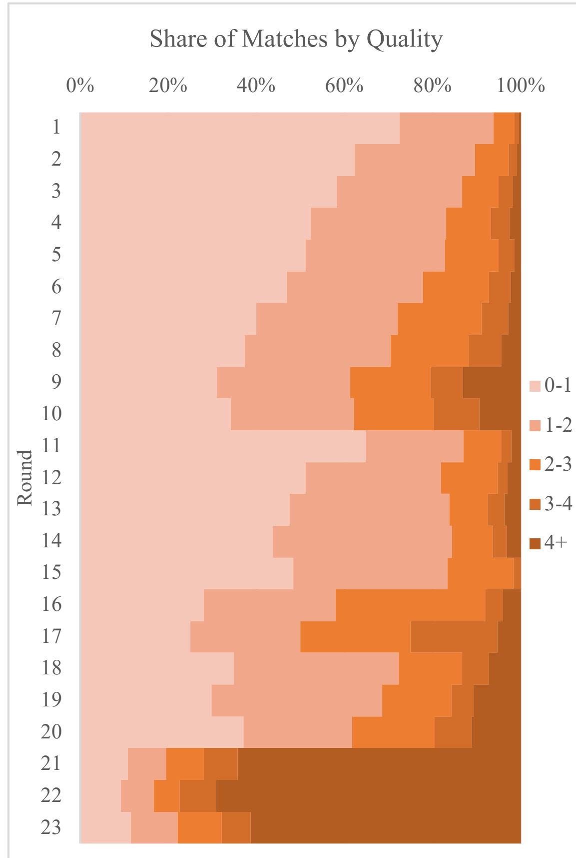
Round	Number of cells	No small cells		With small cells	
		CE Interview	CE Diary	CE Interview	CE Diary
1	1502	47.16	46.11	74.52	70.31
2	1381	7.88	8.49	5.09	6.06
3	1238	3.94	4.07	2.81	3.38
4	1084	4.23	4.30	1.77	2.05
5	930	3.30	3.18	1.50	1.79
6	791	4.01	4.07	1.26	1.79
7	635	3.84	3.74	1.11	1.30
8	476	3.94	3.58	1.31	1.62
9	318	6.28	5.97	1.82	2.12
10	160	6.20	6.16	2.35	2.34
11	400	2.00	2.16	1.38	1.57
12	360	0.92	0.77	0.26	0.32
13	320	0.42	0.52	0.14	0.19
14	280	0.36	0.38	0.15	0.15
15	240	0.25	0.35	0.07	0.14
16	200	0.32	0.38	0.10	0.12
17	160	0.17	0.27	0.12	0.13
18	120	0.33	0.40	0.16	0.20
19	80	0.25	0.49	0.13	0.29
20	40	0.52	0.59	0.39	0.37
21	140	1.72	1.89	1.60	1.74
22	20	0.70	0.91	0.70	0.70
23	4	1.28	1.23	1.28	1.31

Given that a substantial proportion of records are matched in the later rounds, a natural question arises: what is the distribution of characteristics of individuals matched in earlier rounds compared to those matched in later rounds in the donor and recipient files? Intuitively, we would expect matches that occur in earlier stages and use more detailed subgroups to be of better quality than matches that occur in later stages. Rather than comparing the distribution of each characteristic by round, we report how similar the characteristics of the statically matched individuals are, using the first principal component as a measure of similarity. Specifically, we construct a distance measure defined as the absolute difference of the first principal component between the recipient and its matched donor observation, i.e., $d = |pc1_d - pc1_r|$. We use this value and construct five groups (0–1, 1–2, 2–3, 3–4, and 4+) representing different levels of matching quality based on this distance measure. The higher the proportion of individuals with a match in the first group (0–1), the better the quality. Figure 3-2 shows the distribution of the grouped distance measure component across rounds for matching the Interview and Diary samples. From lighter to darker bars, each bar represents the share of a particular group in the total number of observations that are matched (x-axis) in each round (y-axis).

Figure 3-2 Matching quality by round in the matching of the CE Interview and CE Diary samples with ATUS (based on the shares of the matching distance in each round)

<table>
<tr><td colspan="2">No small cells</td></tr>
<tr><td>CE Interview</td><td>CE Diary</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

With small cells

CE Interview	CE Diary
	

As expected, we observe that the matching quality in the first round is the highest. About 72 percent (75 percent) of the matched cases in round one with (without) small cells occur between individuals with a distance lower than 1. As the number of rounds increase, however, we observe a decline in the quality of the matches. After match 21, most matches are low-quality, with over 60 percent being matched to donors with a distance of 4 or above.

Interestingly, there are no notable differences in the distribution of observations based on match quality across rounds when comparing the Interview with the Diary match. Furthermore, we do not see significant differences in matching quality between the procedure that allow for small cells and the one that restricts the use of microcells. That being said, because there is a higher proportion of individuals matched in the first round when allowing for micro-cells, there is a slight advantage in terms of matching distance or matching quality compared to the case with restrictions. However, the small gains in matching quality do not justify the drawbacks associated with potential model overfitting. Because of this, we implement the statistical matching procedure that restricts cell sizes.

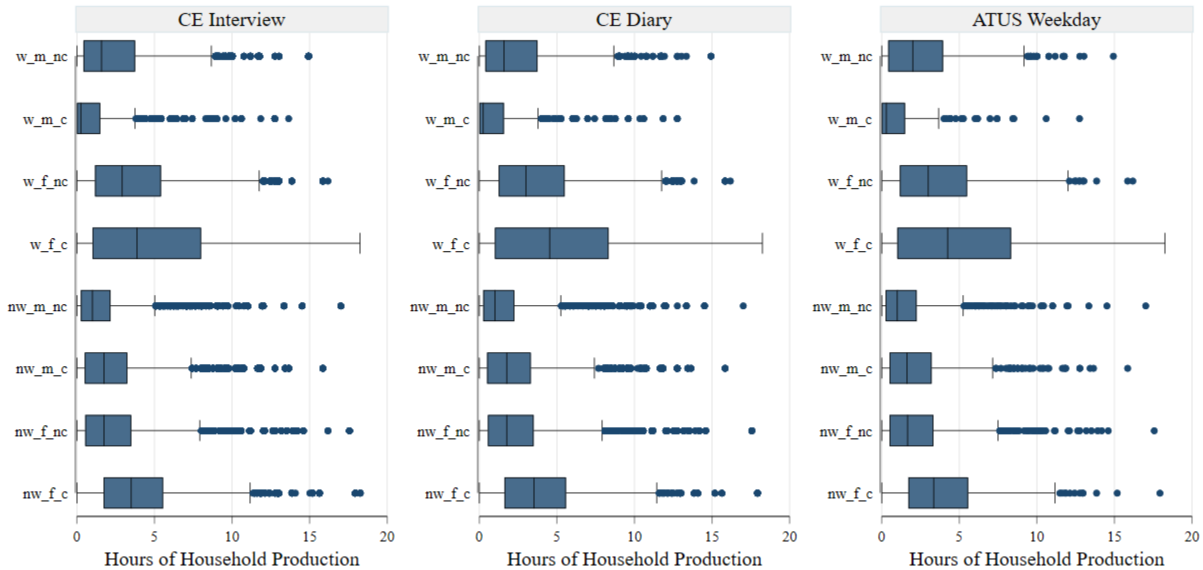
3.3.1.2.3 Quality assessment

The usual practice in assessing quality is to check if the marginal distributions have been transferred from the donor to the recipient file. The verification involves comparing the imputed distribution and actual distribution (e.g., distribution of imputed hours spent on household production by employed women in the recipient file with the distribution of imputed hours of the same group in the donor file). In addition, comparisons of subgroup means, medians, and standard deviations are made across the two files for the key population subgroups that may figure in the potential uses of the imputed data. We present below the results from following this analysis path for the double matching procedure (i.e., treating the ATUS weekday and weekend diaries as separate donor files) without using small cells.

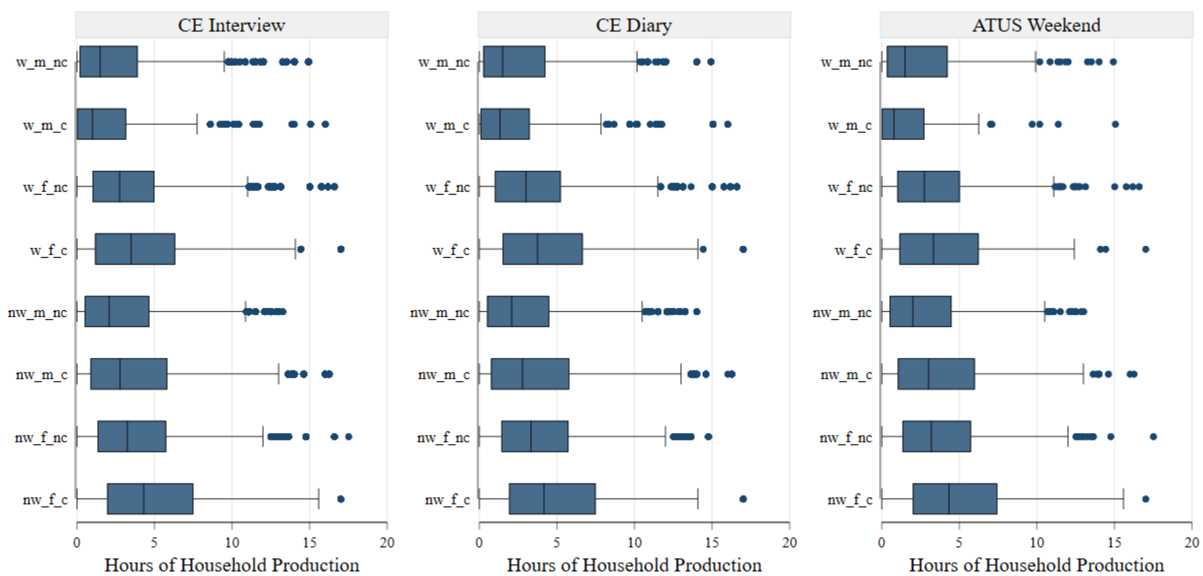
We begin with the results for the eight subgroups formed by three key variables shaping the time spent on household production: gender, presence of children, and employment status (employed or not). A feature of the constrained statistical matching procedure is that it would generally replicate the distribution of the variable of interest with near perfection in the recipient file. This feature is illustrated in Figure 3-3, which shows the distribution of daily hours of household production by subgroup in the donor and recipient files.

Figure 3-3 Box Plots: Distribution of daily hours of household production on weekdays and weekends, imputed vs. observed, statistical matching (double matching method without small cells)

A. Weekdays



B. Weekends

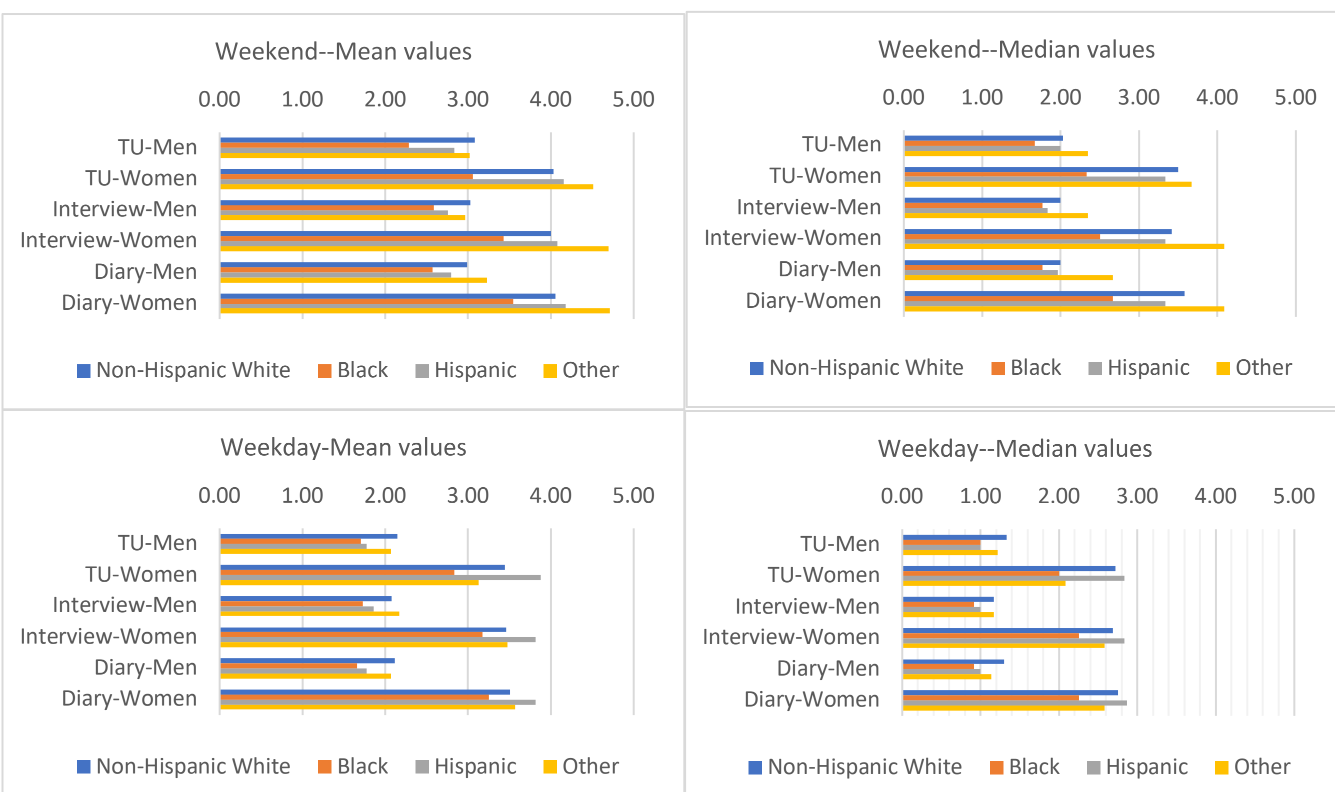


Notes: 1. (n)w=(Not) Employed; m=Men, f=Women; (n)c = (No) Children. 2. "Hours of household production" refers to our preferred definition of the time spent on household production, the sum of time spent on cooking, other housework and shopping, direct childcare, supervisory childcare, and adult care.

We next examine how close the imputed and observed mean and median values are for key population subgroups differentiated by gender. Ideally, we would like the imputations to replicate the same relationship we observed in the ATUS between a specific characteristic and number of hours (e.g., the average hours for women tends to increase with the number of children in the household). This target of imputation would be met automatically if the imputed and "true" (ATUS) values coincided. Our degree of success in meeting the target will be higher the closer the imputed values are to the ATUS values of means as well as medians.

We analyzed statistics for 33 subgroups, each differentiated by gender.[23] The subgroups were created based on the following characteristics, taken one at a time: race or ethnicity (four groups), presence of children five years or younger (two groups), employment status and presence of children in the household (four groups), number of children in the home (four groups), number of adults in the household (three groups), age (five groups), family type (two groups), education (four groups), and family income bracket (five groups). The imputed and observed values were separately compared for the weekday and weekend matches for the Interview and Diary samples. Our online data appendix reports detailed estimates. Here, we provide a qualitative overview.

Figure 3-4 Actual (ATUS) values compared to imputed values, by sex and race/ethnicity (daily hours)



To illustrate our assessment procedure, consider the results for racial/ethnic categories shown in Figure 3-4. Let us first look at the mean values of daily hours spent on weekend days by men in the ATUS (labeled "TU" in the figure). We observe that there is only a negligible difference (less than 15 minutes per day)[24] between non-Hispanic Whites, those in the "other" group (a residual racial category), and Hispanics. On the other hand, the biggest gap is between non-Hispanic Whites and Blacks: the former spends about 50 minutes more per day, on average, than the latter. Our imputation for the Interview and Diary samples has preserved the same pattern in the weekend averages for men, i.e., small gaps between groups other than Blacks and a substantial gap between non-Hispanic Whites and Blacks. A quick inspection of the weekend averages for women in the ATUS would show a similar pattern as those for men, except that the largest gap is between women in the "other" group and Black women (a difference of 88 minutes per day). Our imputations reproduce the same pattern in the Interview and Diary samples. Weekday averages show no sizeable differences among subgroups in the ATUS, and our imputations reflect the same.

To complete the picture regarding racial differences, we should look at the value for the average person in a subgroup, i.e., the median, in addition to the average value for the subgroup. Estimates for men from the ATUS show that the median weekend values of groups other than Blacks display trivial differences from each other; the largest gap is between Blacks and those in the "other" group—the typical person in the latter group spends about 41 minutes more per day on household production than the former. Our imputations in the Interview and Diary sample indicate an identical pattern. Turning to weekday median values for women in the ATUS, we find the same results as for men, though the gap in median value between Blacks and "others" is much higher at 80 minutes. As for weekday median values, we find no notable variations among male racial subgroups, and our imputations reflect this.

In contrast, among female racial subgroups in the ATUS, we find that the weekly median values for non-Hispanic Whites and Hispanics are practically the same and considerably higher than the median values for the other two subgroups, which, in turn, are almost the same. Our imputations do preserve these differences. However, we do end up overstating the median value for the "other" and, therefore, understating the gap in the median between them, on the one hand, and non-Hispanic Whites and Hispanics, on the other.[25]

[24] If we set aside 30 hours in the 48-hour period of the weekend for sleep, personal care, and leisure there are 18 hours left. A difference of 0.5 hours would thus make up less than 3 percent of the maximum time available for household production for those who are not at their jobs on weekends.

[25] It may be noted that the "other" group constitutes only about 7 percent of the female population under study here. The gap between non-Hispanic White and "other" women was respectively 38, 6, and 10 minutes per day in the

Similar results were also, in general, found for other subgroups. Overall, our results indicate that the matches replicated the ordering of group mean values formed by each characteristic. The same general observations also hold for the median values. We summarize the picture in Table 3-6. Panels A and B report our weekend and weekday matches assessment, respectively. If the matching could replicate the same qualitative relationship between the subgroups and time (as we saw above for the racial/ethnic characteristic), we have flagged the characteristic as "Yes."

On the other hand, if our assessment raised any concern regarding the replication, we indicated "No." There are 72 cases in each panel. We marked "No" for the weekend matches in seven instances. One more problematic case was found for the weekday matches. As it turns out, the problematic cases pertained to the same characteristics in both matches: the number of adults in the household and education. Also, in both instances, we noted issues in matching records for women rather than men.

Table 3-6 Classification of characteristics based on whether the relationship between the characteristic and time is reflected in the matching imputation ("Yes") or not ("No"), by type of statistic, sample, and sex

A. Weekend matches

	Mean				Median			
	Interview		Diary		Interview		Diary	
Characteristic	Men	Women	Men	Women	Men	Women	Men	Women
Race/ethnicity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Presence of young (<=5yrs) children in household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Employment status and presence of children in the household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of children in the household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of adults in the household	Yes	No	Yes	No	Yes	No	Yes	No
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	No	Yes	Yes	Yes	No	Yes	No
Family income bracket	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

ATUS Interview and Diary samples. In our judgement, the size of the group and gaps are such (i.e., relatively small) that no major concern needs to be placed on the imputations with regard to racial groups.

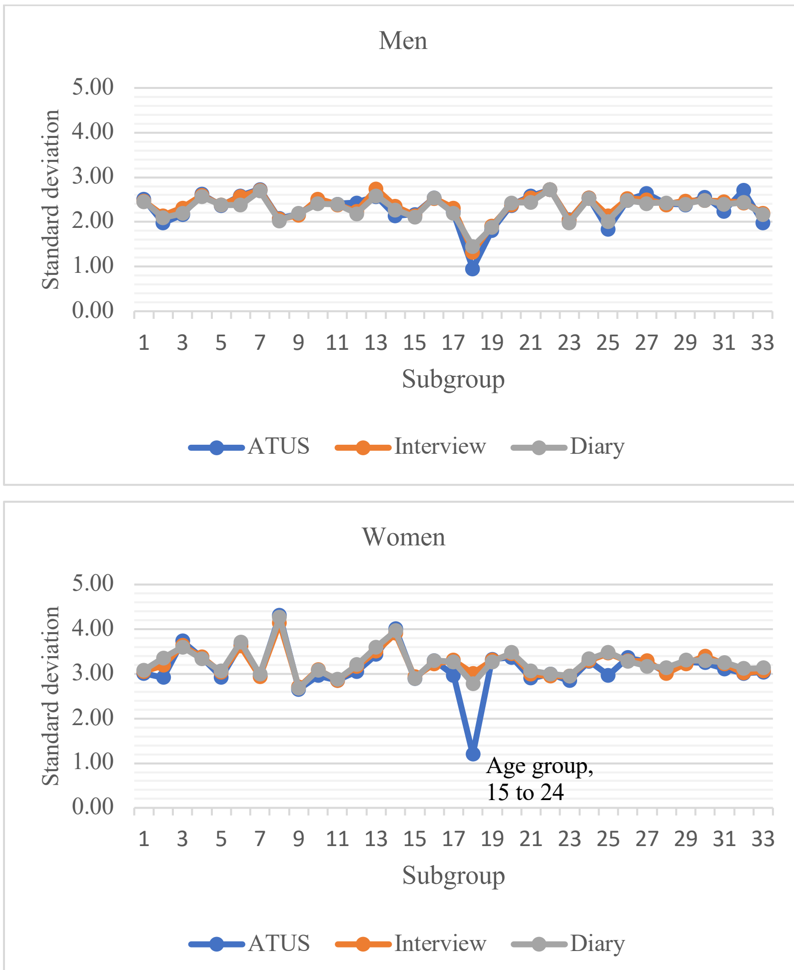
B. Weekday matches

Characteristic	Mean				Median			
	Interview		Diary		Interview		Diary	
	Men	Women	Men	Women	Men	Women	Men	Women
Race/ethnicity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Presence of young (<=5yrs) children in household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Employment status and presence of children in the household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of children in the household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of adults in the household	Yes	No	Yes	No	Yes	No	Yes	No
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	No	Yes	No	Yes	No	Yes	No
Family income bracket	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Concerning the number of adults in the household, the ATUS shows that there is practically no difference between the mean and median values of time for women in one-adult households vs. three-plus–adult households. However, the values are higher for women in households with three-plus adults in our imputation. For the weekend matches, the gap ranges from 7 to 15 minutes per day, which may be considered small. The corresponding range for the weekday matches is somewhat higher: 20 to 31 minutes per day. The gap arises mainly because our matching overstates the values for women in households with more adults. In our view, the discrepancy is not a matter of concern because of the size of the gap and the fact that women in households with three-plus adults constitute less than 10 percent of the matched records.

We have also flagged the relationship between education and time spent by women on household production. Here, the gradient of time with respect to educational attainment is the same in both “true” and imputed data. The problem is with capturing, in our imputations, the extent of the increase revealed by the ATUS data in the time spent by women with a high school degree relative to women without. We found the biggest discrepancy in the gap in the average weekday values. According to the ATUS, women with a high school degree spend 65 minutes per day more than women without a high school degree, while the gap is only 10 and 8 minutes in our imputation for the Interview and Diary samples, respectively. Our imputation also failed to capture fully the extent of the decline in the time spent by women with some college education relative to women with a high school degree. Here the largest discrepancy occurs for the gap in weekday median values: the ATUS median value for the former group is about 34 minutes (per day) lower than the latter group, and our imputation for the Interview sample shows almost no difference in the median values. Therefore, we advise some caution in interpreting the subgroup differences among women in the categories of educational attainment mentioned above. However, the notably higher amount of time spent by college-graduate women compared to that of women without a college degree shown in the ATUS data is also replicated in our imputations.

Figure 3-5 ATUS weekday standard deviation compared to standard deviation in the matched weekday samples, by sex, subgroup, and sample



Finally, we also compared the dispersion of hours in the imputed data to the ATUS data for the 33 subgroups (discussed above), differentiated by sex. We found that, in general, the imputed standard deviations were very close to the ATUS standard deviations. The closeness was true for the weekday and weekend matches conducted for the Interview and Diary samples, respectively. Since the patterns are the same for weekday and weekend matches, we have shown only the results for weekday matches in Figure 3-5 above. In addition, we can observe that the imputed standard deviation is much higher than the actual standard deviation for women in the youngest age group (15 to 24 years). This group, on average, engages much less frequently with household production activities compared to older age groups, as we would expect. In the remaining 32 subgroups, no such discrepancy is visible, thus fostering considerable confidence in the ability of our imputation to capture the within-subgroup dispersion in the time spent on household production.

3.3.1.3 Imputation by prediction

3.3.1.3.1 Procedures

Our deployment of prediction methods also followed, just as with statistical matching, the treatment of the ATUS weekday and weekend diaries as separate donor files. Weekly data can then be constructed by obtaining a weighted sum of these two imputed values (5*weekday+2*weekend).

As distinct from the matching procedure, the prediction methods require a clear a priori definition of dependent variables to be imputed. Since most potential users of the CE data product would like to have some level of disaggregation of household production activities for their analysis, we decided not to impute a single variable that represents the sum of time spent on various household production activities. Instead, we decided to create different components of household production that were described earlier in connection with our valuation schema: cooking; other housework (including shopping); adult care; direct childcare; supervisory childcare that overlaps with cooking, other housework, or adult care; and supervisory childcare that does not overlap with those activities (see Chapter 2). The time spent by the individual on each disaggregated category was considered a distinct dependent variable.

Ideally, we would like to estimate separate models for key population subgroups. But, the number of subgroups we can effectively use is limited by the available number of observations in the ATUS. This restricts the flexibility in specifying models. Nonlinear models, such as the tobit or interval regression model used here, are more sensitive to the problem of near-collinearity among regressors in a model. Typically, this would manifest as a failure of the estimation algorithm to converge after a reasonable number of iterations. After several experiments, we estimated a multivariate model with all the dependent variables specified in the previous paragraph for individuals in households with children. For individuals in households without children, we estimate a multivariate model without the time spent on childcare and supervisory childcare.[26] The covariates used in the models are the same used in the matching procedure described earlier (see Table 3-4 above). However, there are a couple of important differences. We

[26] A very small proportion of individuals in households without children report spending time on the care of household children. For the purposes of the estimation, we included such individuals in the group of individuals in households with children.

included gender interacted with age, and race as explanatory variables in the models. Also, variables of co-residence with children and number of children were used only in models of individuals with children.

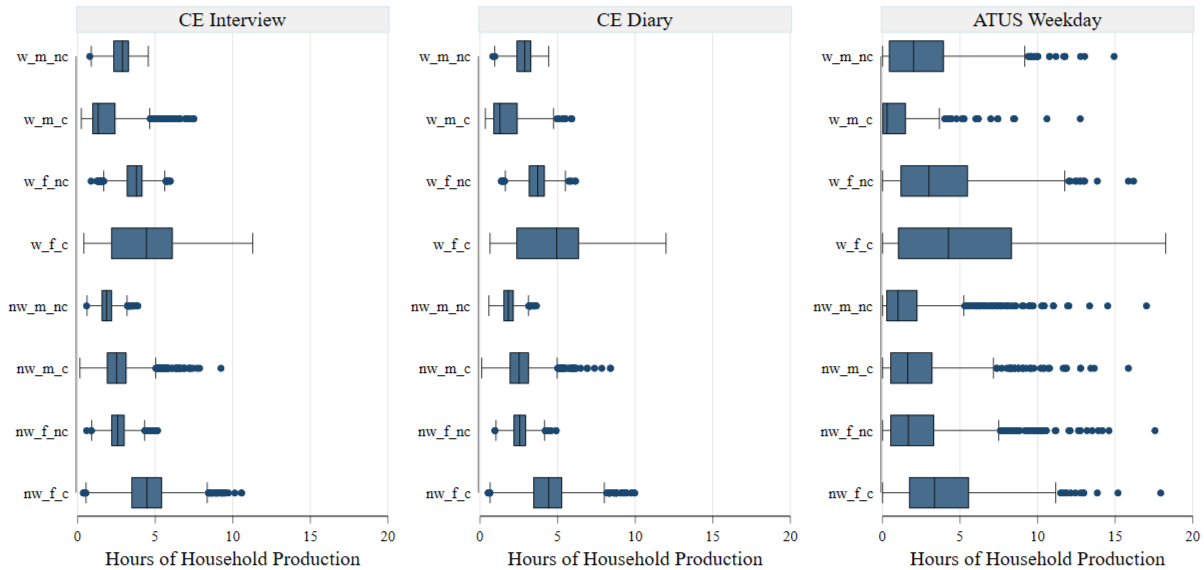
3.3.1.3.2 Quality of imputation

3.3.1.3.2.1 Regression prediction

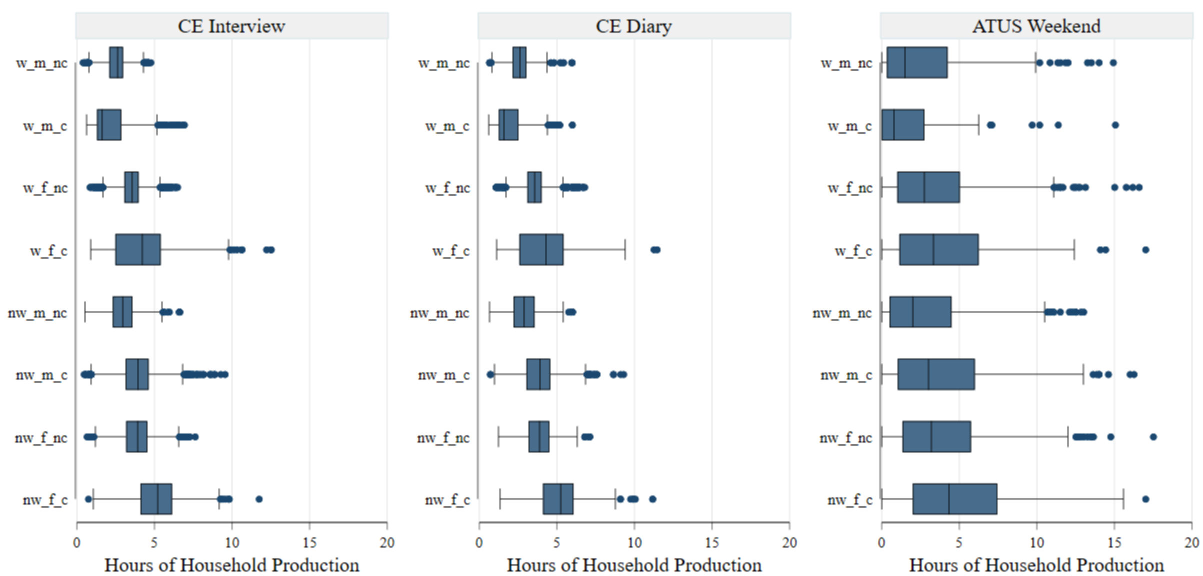
Earlier in this chapter, we described the two variants of the imputation-by-prediction method that we implemented. We begin with the results of the more straightforward approach of regression prediction (RP). We expect RP to perform poorly in replicating the distribution of the hours observed in the ATUS data. This is borne out in Figure 3-6, which depicts the distribution in eight key subgroups formed by interacting gender, presence of children, and employment status (employed or not) in the ATUS and imputed CE data. In contrast, we showed earlier that our statistical-matching procedure carried over the distribution almost perfectly for the same subgroups (Figure 3-3).

Figure 3-6 Box Plots: Distribution of daily hours of household production on weekdays and weekends, imputed vs. observed, imputation by prediction (regression prediction)

A. Weekdays



B. Weekends

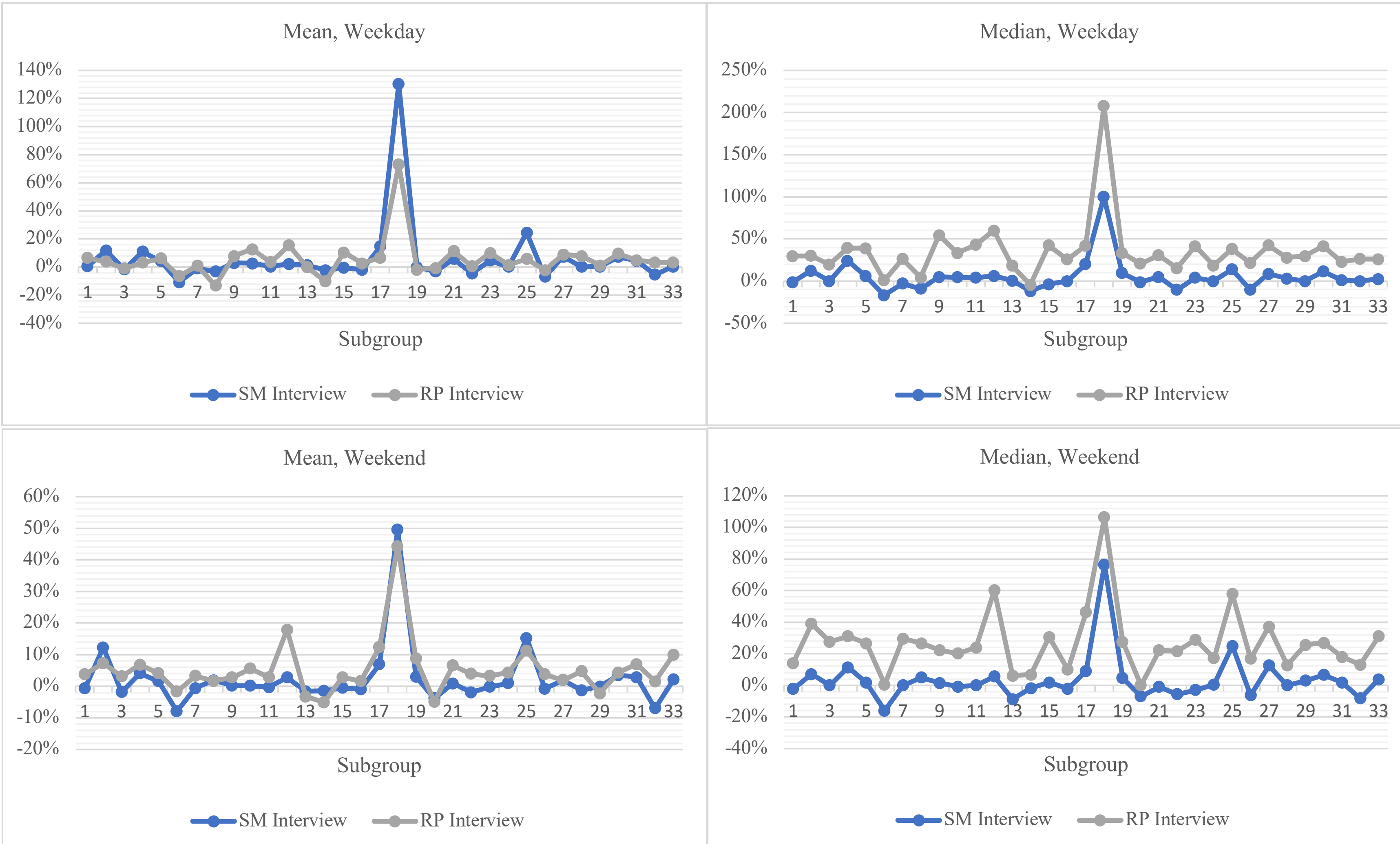


Notes: 1. (n)w=(Not) Employed; m=Men, f=Women; (n)c = (No) Children. 2. “Hours of household production” refers to our preferred definition of the time spent on household production, the sum time spent on cooking, other housework and shopping, direct childcare, supervisory childcare, and adult care.

We next turn to inspect the quality of the RP imputation with regard to reproducing the mean and median values of household production observed in the ATUS data. It may be recalled that our assessment of the results of the statistical matching showed that, in general, matching was able to replicate the qualitative relationships between characteristics and time spent on household production. This ability was based on the relative closeness of the imputed and observed values of two statistics—mean and median—for 33 subgroups, each differentiated by gender. Therefore, we focus on the closeness produced by RP vs. statistical matching (SM) to assess RP. Conducting the exercise for the Interview and Diary samples had similar results. Hence, we report here only the results of the Interview sample. The estimates for the 33 subgroups of women are shown in Figure 3-7 below.[27]

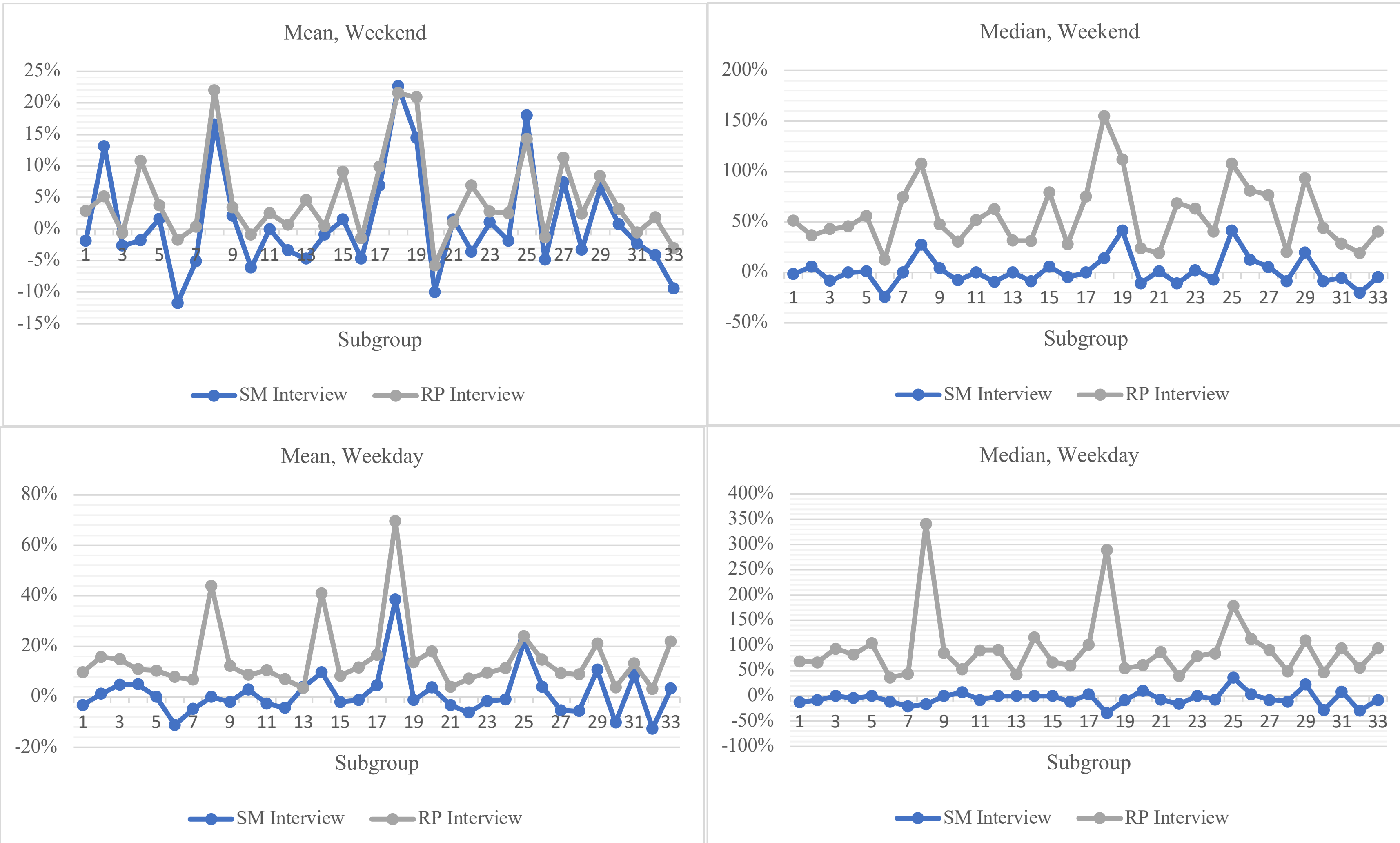
[27] As we noted above, the imputation quality for the 15–24 age group is expected to be poor (see the paragraph below Figure 3-5 above). The spike in the middle of each panel in the figure below is associated with this age group.

Figure 3-7 Percentage deviation of the imputed CE Interview sample statistic from the ATUS statistic: comparison of statistical matching (SM) and regression-prediction (RP) for subgroups of women, by statistic and diary day



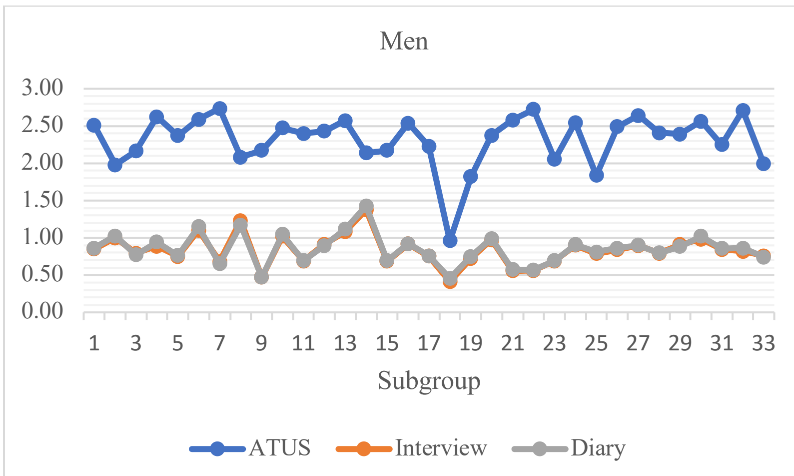
Both methods perform equally well in replicating the ATUS mean values of time spent on household production by women in various subgroups, with one or two exceptions. Deviations are generally bound in the interval of ± 10 percent. The tallest spike occurs for the youngest group of women (between 15 and 24 years of age), which, as we noted earlier, constitutes a relatively small percentage of the female population under study here. However, the SM method outperforms the RP method by a large margin in reproducing the median values. Most of the median values imputed by the SM method fall within a ± 10 percent range of ATUS median values. The percentage deviation spanned by the RP method is much wider.

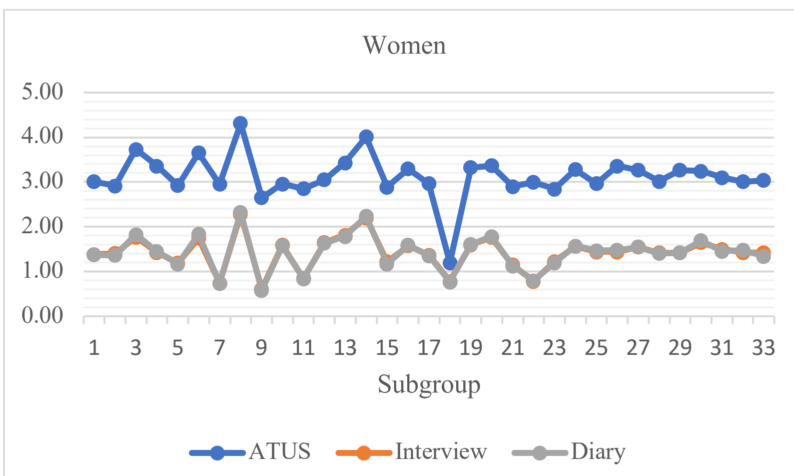
Figure 3-8 Percentage deviation of the imputed CE Interview sample statistic from the ATUS statistic: comparison of statistical matching (SM) and regression-prediction (RP) for subgroups of men, by statistic and diary day



Estimates for the same subgroups of men are shown in Figure 3-8 above. Similar to our findings for the female subgroups, the performance of the RP method is considerably inferior to that of the SM method in replicating the median values for male subgroups. In addition, in several cases, even the weekday mean values for subgroups of men imputed by the RP method exceed the ATUS values by 20 percent or more. This finding contrasts with our assessment of women, where both methods produced similar results for weekday mean values. Statistical matching performed notably better than RP in creating imputed weekday mean values close to the ATUS counterparts for male subgroups and equally well as it did for female subgroups.[28]

Figure 3-9 ATUS weekday standard deviation compared to imputed weekday standard deviation in the CE samples using regression-prediction (RP method), by sex, subgroup, and sample





[28] The comparatively poor performance of RP with respect to capturing the mean values of male subgroups may be due to the larger number of zeroes in the dependent variable for particular subgroups. Also, our RP is not based on a linear regression, where unconditional mean is equal to the predicted unconditional mean by construction. It is, however, hard to pin down the reason for each subgroup because the imputations are done for the sample as a whole.

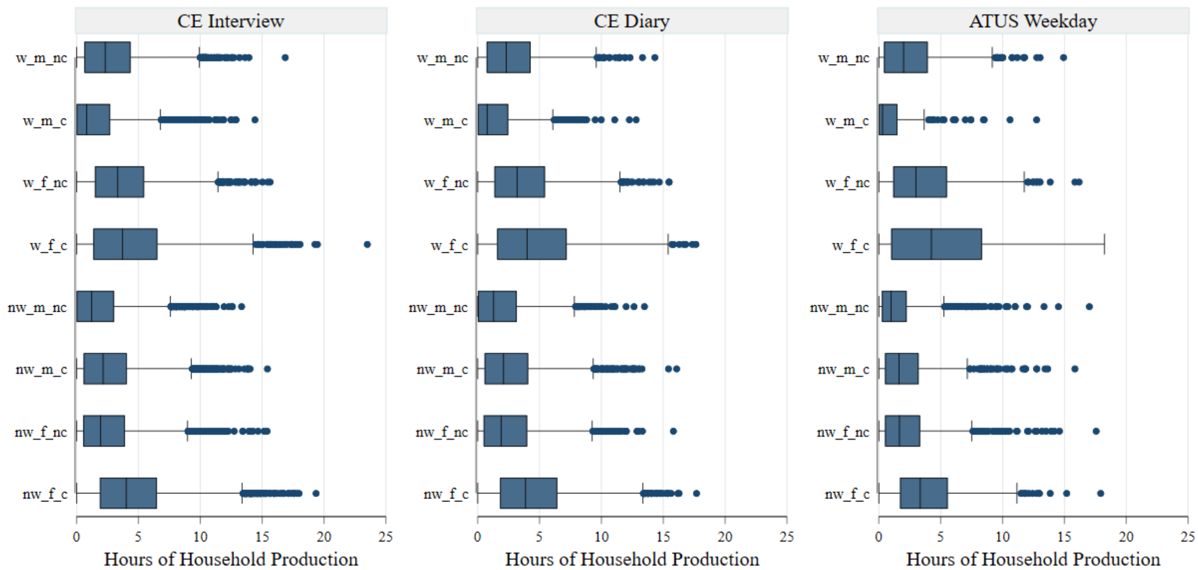
The explanation for the poor performance of RP in adequately replicating the median values lies in its weakness in capturing the within-subgroup distributional properties. By design, the technique aims to capture the conditional means of hours of subgroups defined by characteristics included as covariates in the regression model. For the subgroups we have chosen to analyze here, we expect the method to understate the dispersion in hours considerably. We found evidence for this in the weekday and weekend imputations. Therefore, we depict the results for the weekday imputations in Figure 3-9 above. A comparison with the corresponding figure for statistical matching (Figure 3-5) clearly demonstrates its superiority relative to RP in capturing the extent of the dispersion in the original data.

3.3.1.3.2.2 Multiple imputation

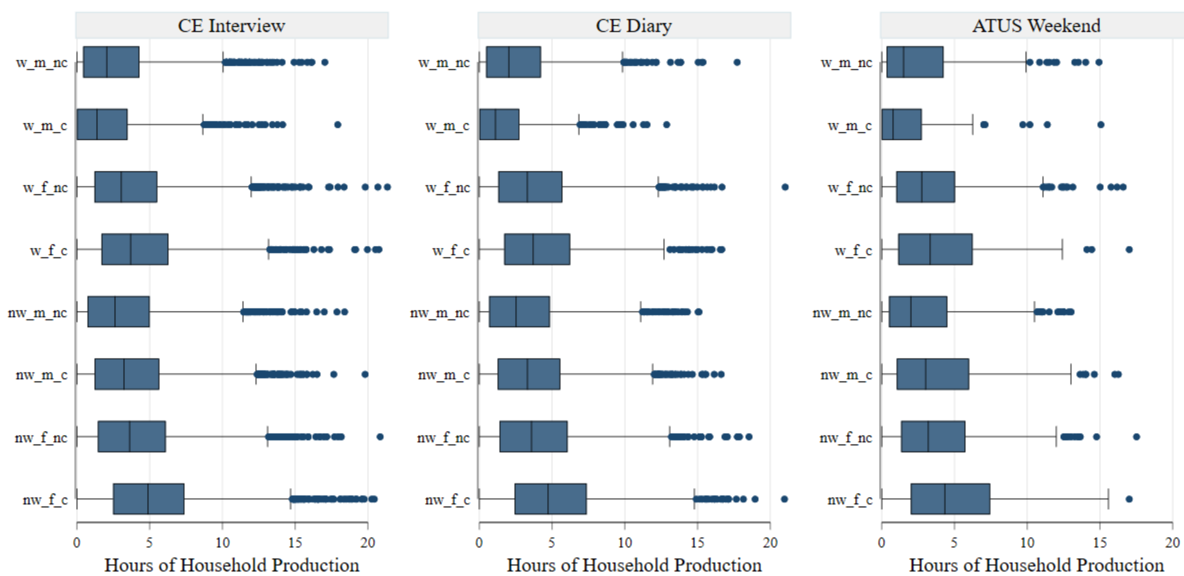
We now turn to the multiple-imputation (MI) variant of the prediction method. As noted earlier, the MI method adds a random component to the predicted value for each observation in the recipient file. The latter is simulated from the estimated variance-covariance matrix of the residuals of the interval-regression model implemented in the donor file. Hence, we expect the MI method to perform better than the RP method in reproducing the distribution of the hours observed in the ATUS within the eight key subgroups (formed by interacting gender, presence of children, and employment status). This is evident by comparing Figure 3-10 below (MI results) with Figure 3-6 (RP results). However, since MI is imputing predicted and SM–actual values, we expect the distribution's tails to be replicated better by SM. Specifically, in some instances, the MI method seems to be creating "too many" large outliers when compared to those observed in the actual (ATUS) data, as can be seen by comparing Figure 3-10 below (MI results) with Figure 3-3 (SM results). This weakness of the MI method is starkly evident for the group "w_f_c" (employed women with children), shown as the fourth group from the top in the figures. The ATUS weekday data shows no outliers (i.e., no observations that fall beyond the right whisker), while the MI method creates quite a few. In contrast, the SM method creates none and is true to the original data.

Figure 3-10 Box Plots: Distribution of daily hours of household production on weekdays and weekends, imputed vs. observed, imputation by prediction (MI method)

A. Weekdays



B. Weekends



Notes: 1. (n)w=(Not) Employed; m=Men, f=Women; (n)c = (No) Children. 2. “Hours of household production” refers to our preferred definition of the time spent on household production, the sum of time spent on cooking, other housework and shopping, direct childcare, supervisory childcare, and adult care. 3. We created five sets of imputed values for each observation using the MI method. Because the quality assessment based on all five showed practically no differences, we report results based on the first set of imputed values.

Since both RP and MI imputations are based on regression models, we do not expect a notable difference in their efficacy imputing the mean values of the hours spent on household production for the 33 subgroups (each differentiated by gender) from the ATUS data. While discussing the RP results, we pointed out that the SM and RP methods performed similarly well in imputing the mean values for women on the weekdays and weekends. This observation also applies to the MI imputation (see Figure 3-11 below).[29] (Because the results from all five sets of imputed values are qualitatively the same, we report only those based on one set of values here.) Similarly, the MI method imputed the mean weekend values for men with reasonable accuracy (Figure 3-12 below). However, as evident in the same figure, MI's performance was weaker for mean weekday values than was SM's. Thus, for the imputation of weekday mean values for men, both variants of imputation by prediction were inferior to SM. As for the imputation of median values, MI displayed the same performance as RP. The imputed values turned out to be higher than the ATUS values by a considerable margin in many subgroups of women in the weekend and weekday imputations. In addition, we found the upward bias to be substantially worse for male subgroups on weekdays and weekends. The upshot is that SM outperforms RP and MI for the imputation of median values.

[29] Because of the qualitative similarity between the assessments of imputations in the Interview and Diary samples, we report here only the results from the Interview sample. The full set of results are available in an online data appendix.

Figure 3-11 Percentage deviation of the imputed CE Interview sample statistic from the ATUS statistic: comparison of statistical matching (SM) and multiple imputation (MI) for subgroups of women, by statistic and diary day

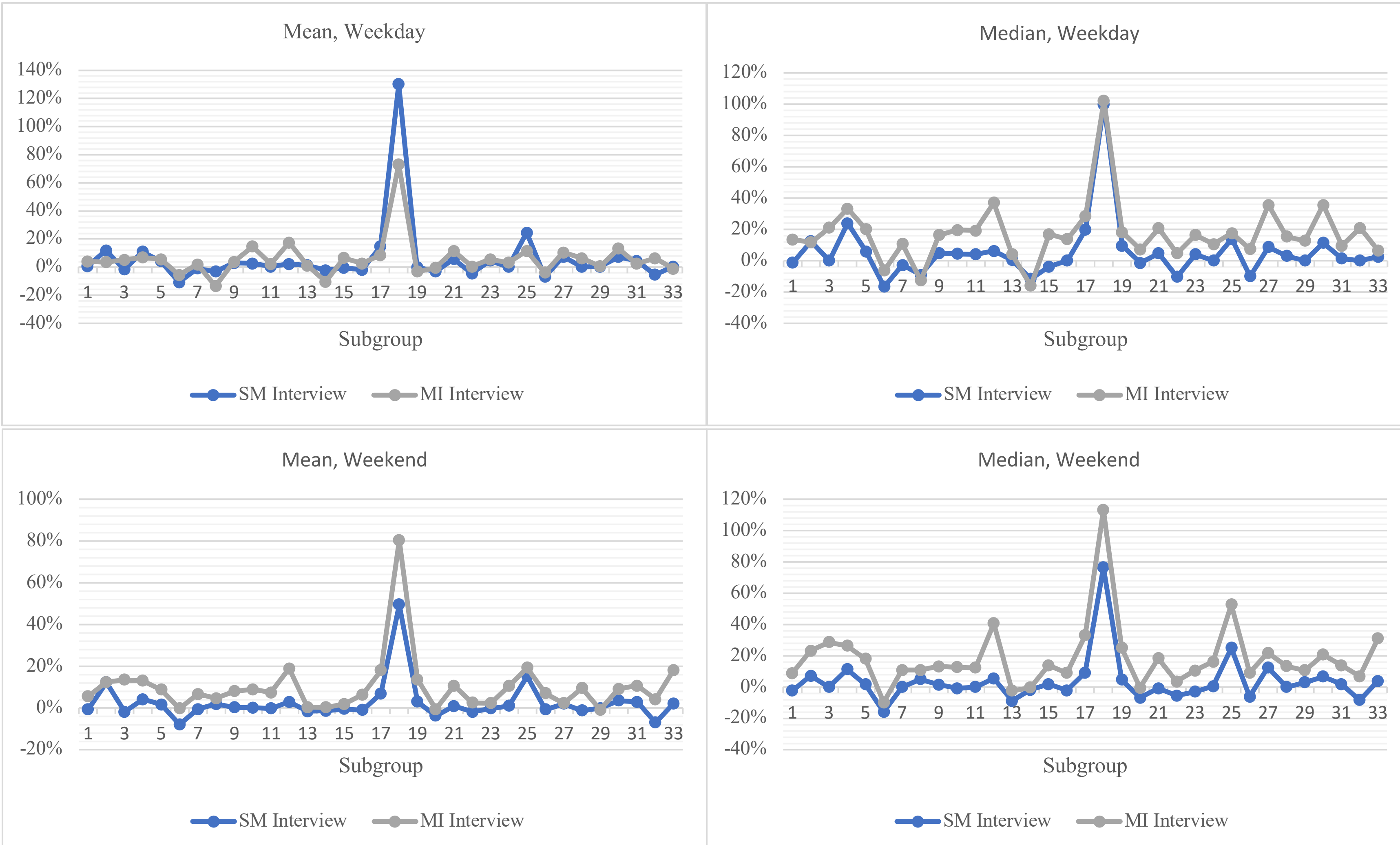


Figure 3-12 Percentage deviation of the imputed CE Interview sample statistic from the ATUS statistic: comparison of statistical matching (SM) and multiple imputation (MI) for subgroups of men, by statistic and diary day

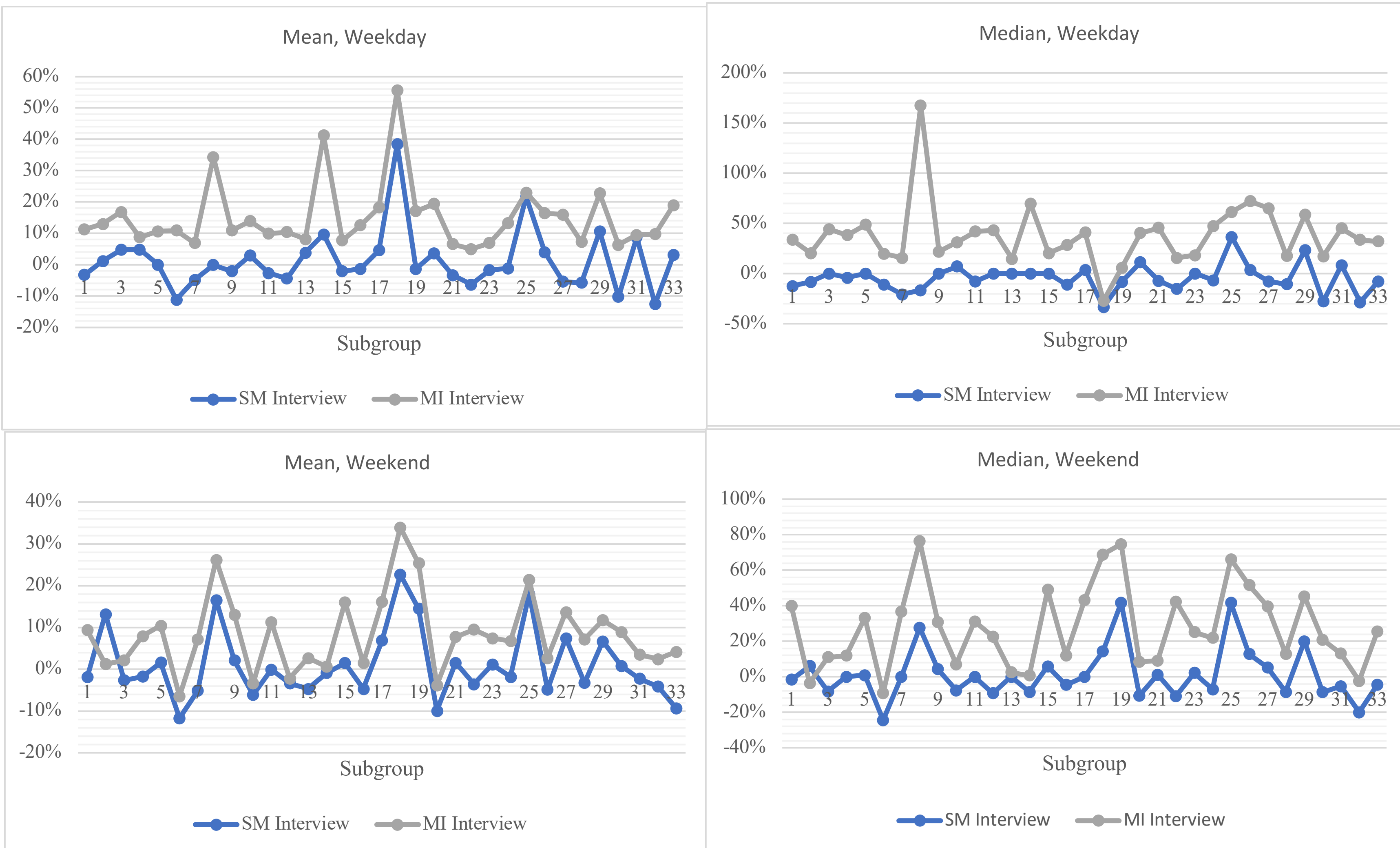
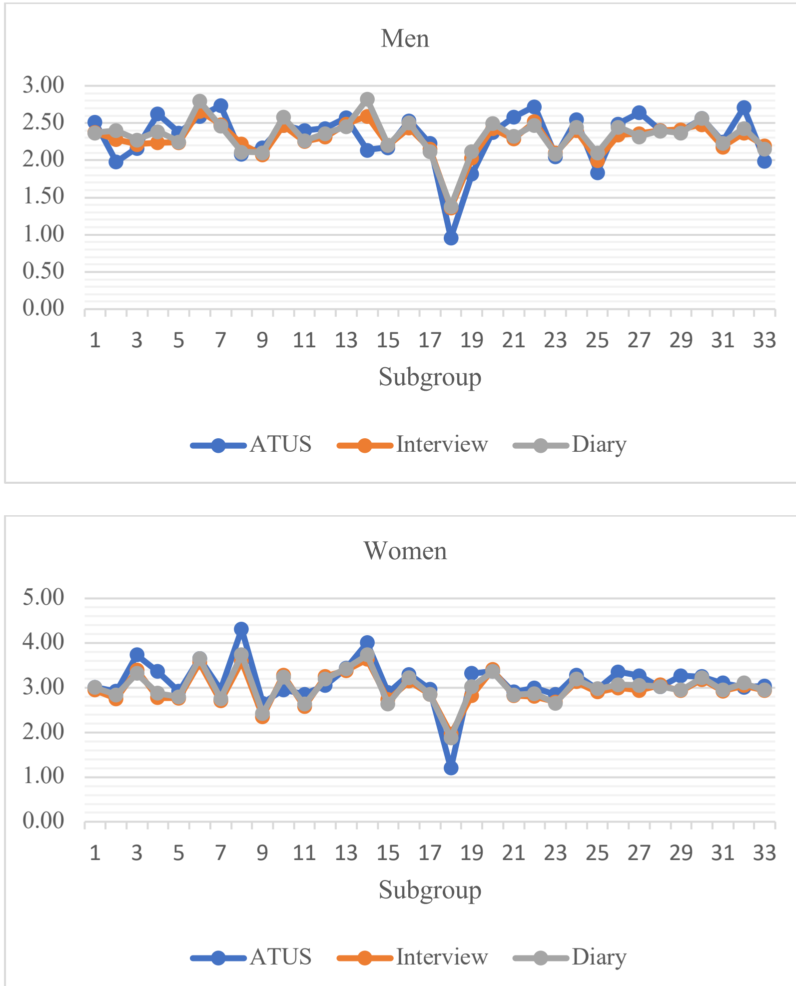


Figure 3-13 ATUS weekday standard deviation compared to imputed weekday standard deviation in the CE samples using multiple imputation (MI method), by sex, subgroup and sample



We expect the MI method to outperform RP in reproducing the dispersion in hours observed in the ATUS samples as a whole and within the 33 subgroups we have been discussing.[30] Indeed, there is no reason to think that the MI method would perform differently than SM in capturing the amount of dispersion. As shown in Figure 3-13 above, unlike the RP method (see Figure 3-9), the standard deviations imputed by the MI method closely track the standard deviations we find in the ATUS and are very much comparable to those generated by SM (see Figure 3-5).

[30] The spike in the middle of each panel corresponds to the 15–24 age group. We outlined earlier why we expect the quality of imputation to be poor for this group (see the paragraph below Figure 3-5 above).

3.4 Aggregation to consumer units

Each of the three methods of imputation assigned time spent on household production for every individual aged 15 years or older in the CE samples. As mentioned, we treat the weekday and weekend diaries as separate donor files. Hence, we impute weekday and weekend values for each recipient in the CE samples. To arrive at the weekly values, we form a weighted sum of the weekday and weekend values with weights of five and two, respectively. This procedure is predicated on the assumption that other weekdays and weekend days (i.e., other than the diary day) are identical to the diary weekday and diary weekend days.[31]

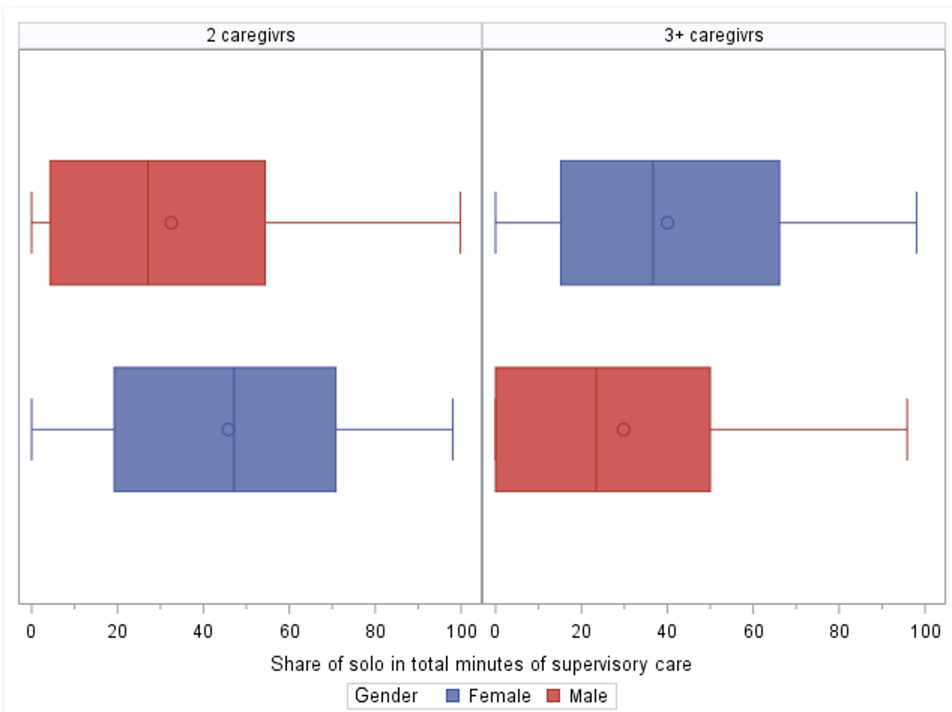
However, we also need estimates of household production (in terms of value and hours) at the consumer-unit or household level. As noted before, since the unit of observation in the ATUS is one person per sample household, we cannot directly ascertain household-level time use. Indeed, the lack of household-level time-use data in the ATUS necessitates the current imputation strategy at the individual level. Consequently, an aggregation step is required in the recipient, i.e., CE files. We add up the imputed times assigned to individuals within the consumer unit to arrive at the household-level aggregates. Our aggregation method assumes that the time spent by an individual on household production is uncorrelated to those of other family members after controlling for the characteristics of the individual (e.g., age), other family members (e.g., employment status of the spouse), and their household (e.g., owner vs. renter). Conceptually, we should not expect the assumption to be problematic in imputing the overall shape of the distribution and its location parameters. This is because we expect the correlation to be low once we account for the relevant covariates. Since the ATUS collects time-use information from only one respondent in each household, we cannot explicitly control for the time spent by other family members on household production in the imputation procedure. However, alternative approaches are possible, and they may lead to different outcomes.[32]

[31] An alternative (more cumbersome) procedure would be to consider time diaries from each day of the week as separate donor files. Given the difference in the number of observations between the ATUS and CE samples, such a procedure would entail, in the case of SM and MI methods, using the same observation from the ATUS for a relatively (relative, that is, to our method) large number of records in the CE samples. Alternatively, when using simple conditional mean imputation, one can consider the imputed data as representing a typical week. With multiple imputation data, the assumption is similar to SM.

[32] For example, an imputation procedure may be deployed in the ATUS to "fill in" the time spent on household production by members other than the respondent in each household. Once the imputations are completed, household-level aggregates of time-use activities could be formed in the ATUS. The household-level ATUS can then be used to impute time-use categories for consumer units in the CE samples. We prefer our method for various reasons: 1) because it uses only the actual information reported by respondents in the ATUS, 2) this alternative approach still requires the conditional independence assumption among household members' activities, and 3) even with data for all household members, matching households is more difficult than matching individuals.

The only exception to this procedure is supervisory childcare. We noted (Chapter 1) that supervisory care may be conducted jointly with other adults in multi-adult households. The incidence of such joint care can be gauged from the "with whom" information ATUS provides for activities, including supervisory care. Utilizing this information, we estimate the share of the supervisory time that the caregiver incurs alone ("solo") and jointly, i.e., another potential caregiver (15-plus years) from the household is present when the respondent is reporting supervisory care (Figure 3-14). Clearly, the typical woman displays a higher share of solo time on household production than the typical man (lines in the box represent the median value). Also, as expected, the median values of solo time are lower in households with two than in those with three or more caregivers.

Figure 3-14 Share of daily supervisory childcare time spent alone (solo) by gender and number of caregivers in the household (percent)



Source: Authors' tabulations from ATUS 2019.

The adjustment we propose can be expressed as follows:

$$\widetilde{s_{ij}} = s_{ij}\alpha_{ij} + s_{ij}(1 - \alpha_{ij})/n_j$$

where $\widetilde{s_{ij}}$ is the adjusted supervisory time of person i in household j, s_{ij} is the actual time on supervisory care, α_{ij} is the share of solo time, and n_j is the number of potential household caregivers present during supervisory care episodes. Thus, for example, if two caregivers report supervisory care of 60 minutes each and they were with each other during 50 percent of that time, each person will be assigned an adjusted supervisory time of 45 minutes, resulting in an adjusted household total of 90 minutes of supervisory time rather than the unadjusted total of 120 minutes.

Table 3-7 Average shares of daily supervisory childcare time spent alone (solo) by gender and number of caregivers (percent)

Gender	Number of potential caregivers	
	Two	Three or more
Female	49	42
Male	33	30

Source: Authors' tabulations from ATUS 2019.

Ideally, we would prefer the individual-specific sharing parameter α_{ij} in our calculations, but that would require our knowing whether the person imputed supervisory time in the CE sample was with another potential caregiver during their episodes of care. In our judgment, reliable individual-level imputations of "with whom" are difficult. Hence, we used the mean values of the sharing parameter (s_{ij}) calculated from the ATUS (Table 3-7). We assumed that the number of potential caregivers (n_j) is equal to to the number of people in the household with positive values of imputed supervisory care time. The appropriate parameter from Table 3-7 was then applied to the imputed hours of each person engaged in supervisory care. In the final step, the adjusted values of individuals were added up to arrive at the consumer-unit level estimate of total supervisory care time.

3.5 Summary

In this chapter, we described the methodologies used and decisions made for imputing time-use data into the CE. Across all imputation methods, we implemented a "double" matching, where weekend and weekday time-use data is imputed separately into the recipient file, aiming to provide a better approximation of a typical week's activity. In addition, when statistical matching is implemented, time-use data is matched individually to each CE quarter, although such a procedure is not needed when using the prediction approach. For the case of statistical matching, specifically, we recommend imposing restrictions on the cell size to avoid problems of overfitting, despite the slight disadvantage the procedure has in matching quality compared to allowing the use of micro-cells.

The three imputation methods presented here have their advantages and limitations, which will depend on the type of analysis one aims to implement. RP is the simplest method to implement. As we have seen, it performs as well as the more complex MI and SM methods in replicating the mean values for most population subgroups we have considered here. However, both prediction methods fail to adequately capture the weekday mean values for male subgroups as SM. Both methods also overstate the median values of several population subgroups by a substantial margin, especially male subgroups. The clear superiority of SM over prediction methods in approximating subgroup median values was evident from our assessment exercise. Of the two prediction methods, RP performs notably worse than MI in predicting subgroup medians. Thus, matching outperforms prediction when replicating location measures for population subgroups.

Turning to the efficacy of capturing the dispersion within subgroups, we find no discernible difference between SM and MI in their imputed values of standard deviations—both are very close to their observed counterparts. Given its logic, the RP method was not expected to perform well in this regard, and our assessment confirmed our expectations. The difference between SM and MI in transferring the distribution seems to be in the ability to mimic the percentile points throughout the distribution. We mentioned earlier the tendency of MI to overstate median values, reflecting its weakness relative to SM in transferring the percentile points. The drawback is also reflected in the relatively sizeable number of high values that appear as outliers in the box plots imputed by MI. SM is the best approach for transferring distributions of time use to the CE, even for smaller subgroups.

The clear disadvantage of the SM and RP methods compared to MI is in the area of statistical inference using imputed values. Before the emergence of MI, most researchers treated the imputed variables in survey data on the same footing as the variables in the survey itself. However, it is now widely known that inference based on such an approach can be misleading because it does not account for the potential errors in imputation. The extent of these discrepancies depends on the details of the statistical model and imputed variables. What seems certain is that the extent cannot be known a priori. There are, thus, tradeoffs in the choice between SM and MI for multivariate analysis using the imputed data, which needs to be explored further.

Our analysis of the imputed data suggests that there are some subgroups for which imputed data is unreliable. Individuals in the youngest age group (15 to 24 years) are one such group, as are individuals in households with three or more children, three or more adults, low levels of education, or low levels of income. Most of these groups are relatively small in size. The general guideline is, therefore, that one should avoid analyzing small population subgroups or subgroups in which a particular type of household production is quite low.

4 Imputing Care Received

Nonmarket-time inputs to providing care for household members may come from people in the household and outside. A novel aspect of our definition of household production is the inclusion of nonmarket care services received by household members from people outside their household (see Chapter 2). Admittedly, the receipt of such services expands the household's consumption possibilities; hence, their imputed value should be included in a broad measure of consumption. The empirical challenge is the lack of direct information. Government statistical agencies in the United States do not collect regular and comprehensive data on the receipt of nonmarket care.

In contrast, the American Time Use Survey (ATUS) presently collects detailed data on time spent by individuals rendering care to nonhousehold children and nonhousehold adults. In the latter case, information is collected on strictly caring activities as well as a broader set of activities encompassing cooking and other housework (including shopping).[33] No information is collected in the ATUS regarding the characteristics of the care recipients or their households.

This chapter describes the sources of data and methods used to impute the value of nonmarket care services received by individual consumer units in the CE Diary and Interview samples. In light of the data limitation noted above and discussed further below, it is evident that our imputations are not as comprehensive here, relative to the imputations performed using the ATUS of the time spent on household production (described in the previous chapter). However, as we shall see later, the contribution made by its members far outweighs that made by persons outside the household in terms of expanding the consumption possibilities of the household. The upshot is that the coarse imputations for the latter may not substantially impact the overall household production estimates. Yet, we include the imputations for the sake of completeness and advocate for better data collection on this front.

4.1 Care received by children

4.1.1 Background

The people outside the household providing nonmarket care services may be family members that live outside the household (e.g., older sibling, step or biological parent), members of the extended family (e.g., grandmothers or aunts), or nonrelatives (e.g., friends or neighbors). Scholars have examined the role of such care in investigating the constraints parents of young children, especially mothers, face in entering

[33] The broader set of activities consists of housework, cooking, and shopping assistance; house and lawn maintenance, and repair assistance; animal and pet care assistance; vehicle and appliance maintenance/repair assistance; financial management assistance; and household management and paperwork assistance.

the labor market and the fertility decisions of people of childbearing age.[34] However, this type of care is mainly categorized under "informal" care, which includes paid care services provided by individuals (e.g., babysitters). For example, a recent study on the relationship between the availability of informal childcare and parental employment during the COVID-19 pandemic in the United States laments the failure of existing research to distinguish between different forms of informal care. Based on data from an online survey, the authors find that, while highly-educated and higher-income families relied more on paid informal care arrangements during the pandemic, those with lower education and income found relatively more support from older children and extended family in maintaining their hours on the job (Zang et al. 2022).

Publicly provided or subsidized formal childcare can go a long way toward meeting the childcare needs of many working parents. Yet, the growth of jobs with non-standard work schedules and irregular or contingent ("on call") hours, as well as the growing share of those with multiple jobs over the last three decades, have made the task of combining care and employment more challenging for many working parents, especially single mothers (Brady 2016; Collins and Carson 1998). There is also the issue that the considerations involved in seeking replacement childcare are, for most people, substantially different than those entering into, say, substituting restaurant-made for homemade meals. A critique of the National Childcare Strategy introduced by the Labour government in the UK in 1997 focused on its exclusive reliance on a regulated market approach and purely financial incentives. The authors' empirical study documented the widespread prevalence of a mix of market and nonmarket childcare—"the childcare jigsaw"—among working parents in an urban area (Tyneside) and a strong preference for nonmarket childcare, especially from grandparents (Wheelock and Jones 2002; see also Casper 1996 for a report on the conditions of employed mothers in the US around the same time).

The Survey of Income and Program Participation (SIPP) has collected information on childcare arrangements from 1985 onwards. In 2019, reference parents[35] of children under 15 were asked questions regarding childcare arrangements during the fall of the reference period. Questions pertained to the "typical" week of the fall season. Only employed or school-going parents were asked. They probed whether the child or children were cared for by parents, siblings older than 14, grandparents, other relatives, nonrelatives, family daycare provider, childcare center, before/after school programs for

[34] A recent example of empirically comparing the impact of formal and informal childcare arrangements in Southern Europe (where provision of childcare by grandmothers is relatively high) on women's employment and fertility is Aassve, Meroni, and Pronzato (2012). On the other hand, Cardia and Ng (2003) provide a general equilibrium treatment of the issue of grandparents' care versus formal care.

[35] Mother is the reference parent in households with both parents; father can give responses only if the mother is not available for an interview. In single-parent families, the parent who lives with the child is the reference parent. If neither parent is in the household, the guardian serves as the reference parent.

children 3 to 14 years) or self (i.e., left alone). Estimates for 2019 indicate that 45.9 percent of children under 14 were under the care of older siblings, grandparents, relatives, or nonrelatives during a typical fall week. Since parents can use multiple sources of non-parental care, the previous proportion may also include children that receive care from centers. On the other hand, only 15.4 percent of children under 15 and 26.2 percent of children under five were exclusively cared for by centers (family daycare plus childcare centers) in 2019.

The SIPP collected more detailed information regarding childcare arrangements before 2012. While only employed mothers were queried between 1985 and 1995, nonemployed mothers were added in 1996. Noteworthy from our perspective is that information on weekly hours spent by preschoolers in each care arrangement was also collected. For example, estimates[36] for 2011 indicate that, on average, preschoolers with employed mothers spent 23 hours, and those with nonemployed mothers spent 15 hours under the care of their grandparents (Laughlin 2013). Further, given our goal of estimating the care received from nonhousehold members, we need to identify whether the caregivers (grandparents, relatives, or nonrelatives) live with the child; otherwise, we would end up double-counting the care received. Earlier rounds of SIPP did permit researchers to distinguish between nonrelative caregivers that lived with the child from those that did not. However, the earlier and present rounds of SIPP do not make that distinction for relatives (older siblings, grandmothers, aunts, etc.). Finally, we also need to know if the care received was directly in return for payment. If that were the case, we would not consider it "nonmarket." Unfortunately, the earlier and current rounds of SIPP do not allow us to implement this distinction because they do not collect childcare expenditures disaggregated by type of care provider.

Perhaps the best available data for alternative care arrangements for children under six are currently collected in the Early Childhood Program Participation Survey (ECPP) module of the National Household Education Survey (NHES). The ECPP has been conducted seven times: 1991, 1995, 2001, 2005, 2012, 2016, and 2019. The latest round coincides with the year we seek estimates (see Cui and Natzke 2021 for a description of the main findings from the latest round and Herbst 2023 for a broader view). The survey collects the incidences of different care arrangements and durations of care. As noted above, the SIPP does not contain any information on the duration of care. Unlike the SIPP, however, the ECCP covers only preschoolers. The classification of care arrangements is similar to the SIPP: parents, relatives, nonrelatives, and centers not located in a private home (including a daycare center, preschool, and pre-kindergarten). As distinct from the SIPP, the survey records the payments for each care arrangement. Estimates indicate that among the children who received care from nonrelatives who did not live in their

[36] Average hours are based on those who reported using that specific arrangement.

household, only a tiny minority (11 percent) received it through nonmarket means (i.e., without payment) in 2019.

In contrast, the majority of children (77 percent) who received care from relatives obtained it in the form of nonmarket services. However, the survey does not enable us to distinguish between caregiving relatives who live and do not live with the child. Therefore, we cannot easily classify the care as rendered by a nonhousehold member. But, we can ascertain whether the relative cared for the child in the child's home, another home, or both places. We may classify those in the latter two categories as nonhousehold members. However, we may be understating the number of nonhousehold caregivers because at least some of those who provide care in the child's home will not be members of the child's household (e.g., the child's grandmother who comes from her home to care for the child while the child's parents are at work). By this reckoning, 71 percent of children who receive care from "nonhousehold relatives" do so without payment. Thus, the ECPP offers much better information than SIPP for estimating nonmarket care received by preschool children from people outside the household.

Before we compare ECPP with the CE samples and outline the imputation procedure, it is helpful to summarize the broad picture from the 2019 round of ECPP. Boys are slightly overrepresented compared with girls among the recipients of nonhousehold, nonmarket care (54 vs. 46 percent). The recipients are also disproportionately non-Hispanic White rather than non-White (58 vs. 42 percent). Children from families where both parents are employed constitute nearly 70 percent of care recipients, while only 45 percent of all children live in such households.

4.1.2 Comparison of ECPP and CE samples

The NHES—based on a nationally representative sample covering 50 states and the District of Columbia—was administered between January and August 2019 by the US Census Bureau. As noted above, the ECPP is a module of the NHES. Questions in the ECPP were posed to the parent or guardian in the sample households regarding *one* child under the age of six and not enrolled in kindergarten. There were 7,076 children in the ECPP module, representing the national population of roughly 21 million preschoolers.

We were attracted to the ECPP because it contains data on the care received by the child from relatives and nonrelatives.[37] For each type of caregiver, information is collected only for regular care arrangements, i.e., scheduled at least once per week. Duration of care given is obtained from responses to questions

[37] The ECPP questionnaire is available at https://nces.ed.gov/nhes/pdf/early/2019_ecpp.pdf

regarding "about how many hours each week does this child receive care from this relative (nonrelative)?"[38] We can assess whether the arrangement is of a nonmarket type by examining the response to the "Yes or No" question: "Is there any charge or fee for the care this child receives from the relative (nonrelative), paid either by you or some other person or agency?" As noted in the previous section, membership in the child's household can be directly gauged from the data for nonrelative caregivers. This is because there is a "Yes or No" question (i.e., "Does this person who cares for this child live in your household?")" regarding nonrelatives. For caregivers that are relatives, we assume that if the respondent answers either "Both" or "Other home" to the question, "Is this care provided in your home or another home?" the relative is not a member of the child's household.

Table 4-1 Demographic composition of the population under six years of age by survey, 2019

		ECPP		CE - Interview		CE - Diary	
		Number ('000)	Percent	Number ('000)	Percent	Number ('000)	Percent
Gender	Female	10,109	48.1	10,489	48.9	9,619	47.8
	Male	10,900	51.9	10,969	51.1	10,519	52.2
Race or Ethnicity	White	10,399	49.5	10,999	51.3	10,575	52.5
	Black	3,323	15.8	3,084	14.4	2,602	12.9
	Hispanic	5,348	25.5	4,964	23.1	4,893	24.3
	Other	1,939	9.2	2,410	11.2	2,067	10.3
Age	Less than one year	4,577	21.8	2,178	10.2	1,819	9.0
	Two years	8,344	39.7	7,894	36.8	7,182	35.7
	3-6 years	8,088	38.5	11,386	53.1	11,137	55.3
Number of children under six years	One	11,173	53.2	11,178	52.1	10,708	53.2
	Two	8,005	38.1	8,118	37.8	7,272	36.1
	Three or more	1,831	8.7	2,161	10.1	2,158	10.7
Number of children, 6 to 18 years	None	11,945	56.9	11,451	53.4	10,851	53.9
	One	5,294	25.2	5,869	27.4	5,410	26.9
	Two	2,743	13.1	2,599	12.1	2,028	10.1
	Three or more	1,027	4.9	1,539	7.2	1,849	9.2
Family type	Couple	16,533	78.7	15,550	72.5	15,160	75.3
	Single mother	3,289	15.7	3,205	14.9	2,865	14.2
	Single father	691	3.3	875	4.1	773	3.8
	Other	496	2.4	1,829	8.5	1,339	6.6

[38] It should be noted that the concept of the duration of caregiving here is different from that employed in the ATUS with respect to care given to nonhousehold children. The latter records the time spent by the respondent in such activity during the previous 24 hours and can include care that is not regularly scheduled, e.g., when the respondent babysits their friends' three-year-old as a random act of kindness so that the friends can go out for a movie. It is arguable that regularly received care is a better measure of care with respect to augmenting a yardstick of consumption or income.

		ECPP		CE - Interview		CE - Diary	
		Number ('000)	Percent	Number ('000)	Percent	Number ('000)	Percent
Parental employment	Both employed	9,508	45.3	11,153	52.0	10,577	52.5
	One employed	9,917	47.2	8,986	41.9	8,861	44.0
	Neither employed	1,584	7.5	1,319	6.1	700	3.5
Education: first parent	Less than HS	2,261	10.8	2,568	12.0	2,495	12.4
	High school	4,200	20.0	4,367	20.4	3,883	19.3
	Some college	5,462	26.0	5,870	27.4	5,681	28.2
	College	9,086	43.2	8,653	40.3	8,079	40.1
Education: second parent	Less than HS	2,170	12.7	2,083	11.7	2,483	14.5
	High school	3,667	21.5	3,902	21.9	4,166	24.3
	Some college	4,217	24.8	4,207	23.6	4,310	25.1
	College	6,978	41.0	7,599	42.7	6,203	36.1
Household income	Less than $30k	3,991	19.0	5,619	26.2	4,621	22.9
	$30k-$60k	4,869	23.2	4,598	21.4	5,124	25.4
	$60k-$100k	4,960	23.6	4,769	22.2	4,343	21.6
	$100k-$150k	3,691	17.6	3,114	14.5	3,533	17.5
	$150k or more	3,498	16.6	3,358	15.7	2,516	12.5
Tenure	Renter	9,060	43.1	8,848	41.2	8,099	40.2
	Owner	11,950	56.9	12,609	58.8	12,039	59.8
Region	Northeast	3,531	16.8	3,563	16.6	3,229	16.0
	South	7,445	35.4	8,050	37.5	7,771	38.6
	Midwest	4,717	22.5	4,931	23.0	4,199	20..9
	West	5,317	25.3	4,914	22.9	4,938	24.5
All		21,009	100	21,458	100	20,137	100

We provide a comparison of the ECPP sample and CE samples in Table 4-1. A notably higher proportion of children less than a year old is seen in the ECPP compared to the CE samples (22 percent vs. approximately 10 percent). Children in all samples are almost identical regarding their gender, race, or ethnicity. There is no substantial difference in terms of household demographic characteristics such as the number of young children (under six years), the number of older children (6 to 18 years), and the type of family (couple, single-female, single-male, or other).[39] There appears to be a lower proportion of families

[39] Definition of the family type had to be different across surveys though the difference is unlikely to matter significantly for the estimates presented here. In the CE samples, we can directly identify only the spouse and child (or grandchild) of the reference person. Hence, our family typology is based on the relationship to the reference person. However, in the ECPP, because the unit of observation is the child, fathers and mothers are identified directly if they live with the child. Father or mother could be biological, adoptive, step, foster, or same-sex partner of parent. An example may illustrate the difference clearly. In the ECPP, if the child's mother lives in the household, we will know who that is, irrespective of the mother's relationship to other people in the household. In the CE, for the same family, if the mother of the child is the niece of the reference person, we will not be able to know for sure the relationship between the child and mother.

where both parents[40] are employed in the ECPP compared to the CE samples and a corresponding overrepresentation of families with only one employed parent.

On the other hand, parental educational attainment (four categories) appears to be broadly similar across samples. We can also observe some differences concerning the distribution of households among the five income brackets. The lowest bracket (annual income under $30,000) is underrepresented in the ECPP relative to the CE samples, while the middle ($60,000–$100,000) and top ($150,000 or more) brackets seem overrepresented, especially when compared to the CE Diary sample.

On the whole, we assess that the samples are aligned well. Therefore, the ECPP can serve as the donor file to estimate the nonmarket care received from nonhousehold people by the consumer units. We next outline the method of imputation and discuss the quality of imputation.

4.1.3 Models and results

For the imputation of care received by household members, we use a modified two-part model to find the probability that a member of the consumer unit will receive nonmarket care from people outside the unit and how many hours of such care they would receive on average. First, the likelihood of receiving care is modeled using a binomial logit model, such that:

$$P(d = 1|x) = \Lambda(x\gamma)$$

The second part models the hours of care received as a Poisson model, using only the subsample that received care. We can express the mean values of hours as

$$E(y|x) = \exp(x\beta)\ \ if\ d = 1$$

For the imputation in the CE, we determine if a member of the consumer unit received any kind of external care using a random draw from a Bernoulli distribution based on the predicted probability $\Lambda(x\hat{\gamma})$. This approach is similar to stochastic imputation but based on a binomial model. It has shown good performance in replicating the recipient individuals' characteristics in the donor sample.

Turning to the imputation of hours received, we use a predicted mean from the Poisson model, which is given by exp $(x\hat{\beta})$. While conditional mean imputation is known for reducing the variation in the imputed data (as shown for other time components), it remains the best approach that minimizes the root-mean-squared error compared to other methods. In addition, the group of households that receive the type of care we seek to impute is relatively small. In the CE samples, households with preschoolers constitute

[40] In line with our previous note regarding family typology, we define parent as the reference person with their own child or grandchild in the household. If the reference person has a married or unmarried partner, that person is defined as the second parent.

about 13 percent of the total number of consumer units. As we show below, the ECPP indicates that preschoolers receiving regular care from kith and kin who are not members of the household constitute under 10 percent of all children under six. Further, our examination of the ECPP microdata showed notable differences among population subgroups of preschoolers in the recipiency rate. These characteristics of the problem suggest that a prediction method with "noise" and a straightforward technique of replicating conditional averages would be sufficient to arrive at reasonably accurate estimates of incidence and duration of care.

We began the imputation procedure by estimating, using the ECPP data, the likelihood of a child under six receiving nonmarket care from outside the household, conditional on the relevant characteristics—derived from categorical variables already outlined in Table 4-1—of the child, their parents, and their household. We list the dummy variables and the estimation results in Table 4-2. Survey weights in the ECPP were used to generate the estimates in the table.

Table 4-2 Estimates of the binary logit model (dependent variable is the recipiency of care by a child under six years)

Variable	Estimate	Standard error
Intercept	-2.468**	0.211
Race/Ethnicity (reference group is non-Hispanic white)		
Black	-0.244	0.264
Hispanic	-0.356**	0.138
Other	-0.448**	0.197
Age (reference group is children under three years)		
3 - 6 years	-0.286**	0.107
Number of kids under six in HH (reference group is one kid)		
Two or more	-0.215*	0.120
Presence of older children (6 to 18 years) in HH (reference group is no older child)		
Older child	-0.361**	0.141
Family type (reference group is a couple- household)		
Not a couple-household	-0.438**	0.186
Parental employment (reference group is households with at least one nonemployed parent)		
Both parents or co-resident single parent employed	1.397**	0.191
Parental education (reference group is those without a college degree)		
Parent 1 college graduate	0.195	0.131
Parent 2 college graduate	-0.222*	0.117
Household income (reference group is under $60,000)		
$60,000-$100,000	-0.177	0.169
$100,000 or more	-0.573**	0.130
Home tenure (reference group is owner)		
Renter	-0.165*	0.096
Region (reference group is Midwest)		
North	0.223	0.186
South	0.007	0.109
West	0.112	0.138

Note: N = 7,061. The unit of observation is a child under six years of age. ** - significant at the 5 percent level, * - significant at the 10 percent level. Unmarked estimates are not statistically significant. Standard errors were calculated using the 80 replicate weights that account for the complex survey design and are available in the data file.

Source: Authors' estimates from the public-use data file of the Early Childhood Program Participation.

Our focus here is not on building a causal explanation based on the model. Still, we highlight a few interesting aspects of the estimates. Some may be worth further scrutiny within a richer analytical and empirical approach, especially considering factors ignored here, such as proximity to kith and kin, involvement in social networks, and preferences.[41] We find that the likelihood of receiving care is lower for Hispanic children and "other" race children than for non-Hispanic White children, holding other covariates constant.[42] The youngest is also favored compared to those in the 3–6 age group. We speculate that the probability of receiving care diminishes with the number of preschoolers in the family. A parent or guardian is more likely to be nonemployed to stay home if there are two or more young children. The need for care from kith and kin outside the household is reduced when there is a stay-at-home parent. The presence of older children (6–18 years of age) seems to reduce the need for external help because the older siblings may engage in caregiving. It is curious that not having two parents in the home reduces the likelihood of receiving care, as we would imagine that being a single parent increases financial and caregiving responsibilities. Perhaps, the estimates indicate that, after controlling for other covariates, such a demand may not be met with a corresponding supply.

The variable reflecting parental employment has an expected positive sign, implying that, if both parents are employed in a couple-household or if the single parent in a single-parent household is employed, they are more likely to receive help compared to families where there is at least one nonemployed (possibly stay-at-home) parent. As proxied by the college graduate dummy, parental educational attainment demonstrates opposite signs for the first (positive) and second (negative) parent. Since we associate greater employment probability with a college degree, we might expect that the child with a college graduate as a second parent may be more likely to receive care. However, the greater preference and ability to afford center-based care may offset this. Finally, income and wealth effects seem to operate in the same direction. Children from higher-income families seem less likely than those from lower-income families to receive care, and renters who generally are less wealthy than homeowners are less likely to get care.

The coefficients from the logit model estimated using the ECPP data were used to predict the probability of receiving care in the CE samples of children under six. We then employed a Bernoulli distribution to assign recipiency to individual children to introduce a degree of randomness. To examine the imputation quality, we provide estimated recipiency in the ECPP and CE samples in Table 4-3.

---

[41] The ECPP contains detailed information regarding the reasons behind the choice of care providers. However, we cannot use that information for the imputation because of the absence of matching information in the CE survey.

[42] This may be due to the higher average number of adults in Hispanic households. See the discussion in Section 5.3.2 below.

Table 4-3 Observed and imputed recipiency of care (percent) among children under six years

Characteristic	Subgroup	Incidence (percent)			Deviation from ECPP (%)	
		ECPP	CE - Interview	CE - Diary	CE - Interview	CE - Diary
Gender	Female	8.8	8.9	8.4	1%	-5%
	Male	9.7	8.9	10.4	-8%	7%
Race or Ethnicity	White	10.9	11.1	10.3	2%	-6%
	Black	8.6	7.1	10.4	-17%	21%
	Hispanic	7.4	7.1	7.6	-4%	3%
	Other	7.0	4.9	8.3	-31%	18%
Age	Under 3 yrs.	10.3	11.2	10.0	9%	-2%
	3 to 6 yrs.	7.7	6.9	9.0	-11%	17%
Number of children under 6 years of age	One	10.1	10.0	10.2	-1%	1%
	Two or more	8.3	7.7	8.5	-7%	3%
Number of children under 6 to 18 years of age	None	10.8	10.2	10.6	-6%	-2%
	One or more	7.2	7.5	8.1	4%	12%
Family type	Not a couple HH	9.6	9.4	9.4	-3%	-2%
	Couple HH	8.0	7.7	9.6	-4%	20%
Parental employment	At least one nonemployed parent	3.8	3.4	3.9	-11%	2%
	No nonemployed parent	13.1	12.0	12.5	-9%	-5%
First parent: college grad?	Not a college grad	8.2	8.6	9.0	4%	9%
	College grad	10.6	9.4	10.1	-12%	-5%
Second parent: college grad?	Not a college grad	9.2	8.7	9.7	-5%	6%
	College grad	9.4	9.3	8.8	-1%	-7%
Income	Less than $60k	8.7	8.4	8.8	-3%	2%
	$60k-$100k	10.7	10.8	12.0	1%	12%
	$100k or more	9.0	8.3	8.6	-8%	-4%
Tenure	Renter	8.2	8.3	9.1	2%	11%
	Owner	10.1	9.3	9.7	-8%	-4%
Region	Northeast	10.9	12.0	12.0	11%	10%
	South	8.5	8.5	9.3	0%	9%
	Midwest	10.2	8.3	9.9	-18%	-3%
	West	8.4	7.9	7.7	-7%	-9%
All		**9.3**	**8.9**	**9.4**	**-4%**	**2%**

Note: ECPP refers to the Early Childhood Program Participation Survey, 2019.

Overall, the imputed incidence in the Interview (8.9 percent) and Diary (9.4 percent) samples are very close to the observed incidence in the ECPP (9.3 percent). The imputation performed reasonably well for the population subgroups considered for our analysis and shown in the table. A simple rule-of-thumb examines whether the imputed and observed recipiencies diverge from each other by more than 15 percent. Only three such instances can be found among the 29 subgroups in the table for the Interview and Diary samples. The first two are for the racial groups “Black” and “Other.” For the Interview sample, the region “Midwest” was the third culprit, while for the Diary sample, it was the “Couple” household. The estimates for these groups should, therefore, be used with caution. For example, for the racial groups, the non-White groups may be combined into a single category for comparison with the White group since the ECPP data indicates that the unconditional averages for the non-White groups are pretty similar. In sum, the imputations faithfully reproduce the differences in incidence apparent in the donor, i.e., ECPP data.

The next step in our imputation procedure was assigning hours of care received by children. We first estimated a Poisson model of hours using the ECPP data on children who received care.[43] The covariates we used here are the same as those in the binary logit model described above. We show the results of the estimation in Table 4-4 below. As with the logit model, survey weights were used in our estimation.

[43] We experimented with a truncated regression model. However, as happens frequently when the number of zeros in the dependent variable is “large” (i.e., the degree of participation is low), the model failed to converge. On the other hand, the Tobit model we estimated faced no convergence problems but the quality of imputation was inferior to the Poisson model we eventually chose.

Table 4-4 Estimates of the Poisson model (dependent variable is the weekly hours of care received by a child under six years)

Variable	Estimate	Standard error
Intercept	-2.468***	0.139
Race/Ethnicity (reference group is non-Hispanic white)		
Black	-0.044	0.127
Hispanic	-0.107	0.083
Other	0.272***	0.089
Age (reference group is children under three years)		
3 - 6 years	-0.007	0.056
Number of kids under 6 in HH (reference group is one kid)		
Two or more	-0.063	0.063
Presence of older children (6 to 18 years) in HH (reference group is no older child)		
Older child	-0.031	0.068
Family type (reference group is couple household)		
Not a couple household	-0.099	0.097
Parental employment (reference group is households with at least one nonemployed parent)		
Both parents or co-resident parent employed	0.392***	0.124
Parental education (reference group is those without a college degree)		
Parent 1 college graduate	-0.136**	0.066
Parent 2 college graduate	-0.056	0.066
Household income (reference group is under $60,000)		
$60,000-$100,000	0.066	0.086
$100,000 or more	0.059	0.096
Home tenure (reference group is owner)		
Renter	0.028	0.072
Region (reference group is Midwest)		
North	0.046	0.090
South	0.176**	0.074
West	0.057	0.081

Note: N = 743. The unit of observation is a child under six years of age. ** - significant at the 5 percent level, *** - significant at the 1 percent level. Unmarked estimates are not statistically significant. Standard errors were calculated using the 80 replicate weights that account for the complex survey design and are available in the data file.

Source: Authors' estimates from the public-use data file of the Early Childhood Program Participation, 2019.

There are some interesting points of difference between the parameter estimates of the hours (Poisson) model and the incidence (logit) model. For example, while the incidence model clearly favors White children, the hours model indicates that, among recipients, children belonging to "other" racial groups receive more hours of care from their kith and kin than White children.[44] The next point of contrast is the role of income and wealth. As noted above, higher amounts of both seem to translate into lower recipiency rates. However, among care recipients, children with higher family income and wealth receive a longer duration of care than children who have low family income or who live in rented homes (presumably, less wealthy on average). It is also notable that having a college graduate as the first parent appears to reduce the duration of the care received by the child in contrast to our earlier finding that it makes the probability of receiving care higher—both compared to children with less educated parents.

[44] It may be recalled that, in the Poisson model, the size of the implied effect can be found by exponentiating the parameter estimates. For example, the parameter estimate for parental employment status is 0.39 and $\exp(0.39) = 1.48$. The average hours of care received by a kid in a family where both parents are employed or a family with a single, employed parent is 1.48 times higher (i.e., approximately 50 percent higher) than the kids in the reference group, i.e., kids with at least one parent who is not employed, other covariates held constant.

Table 4-5 Observed and imputed weekly hours of care received by children under six years

Characteristic	Subgroup	Weekly hours			Deviation from ECPP (%)	
		ECPP	CE - Interview	CE - Diary	CE - Interview	CE - Diary
Gender	Female	19.8	19.8	19.7	-2%	0%
	Male	19.0	19.8	20.0	1%	5%
Race or Ethnicity	White	19.5	19.9	19.6	-1%	0%
	Black	18.1	19.5	19.0	1%	5%
	Hispanic	18.1	17.6	18.6	-2%	3%
	Other	24.2	25.3	25.8	7%	7%
Age	Under 3 yrs.	19.4	19.9	19.5	0%	0%
	3 to 6 yrs.	19.2	19.5	20.2	0%	5%
Number of children under 6 years of age	One	20.3	20.5	20.6	-1%	1%
	Two or more	18.1	18.7	18.9	1%	5%
Number of children under 6 to 18 years of age	None	19.6	19.8	20.1	0%	3%
	One or more	18.9	19.7	19.4	0%	3%
Family type	Not a couple HH	19.6	20.0	19.7	-1%	0%
	Couple HH	18.1	19.1	20.3	3%	12%
Parental employment	At least one nonemployed parent	14.2	13.9	14.6	-4%	3%
	No nonemployed parent	20.4	20.7	20.8	-1%	2%
First parent: college grad?	Not a college grad	19.8	20.5	20.5	1%	4%
	College grad	18.9	18.8	19.0	-3%	0%
Second parent: college grad?	Not a college grad	19.5	19.9	19.9	0%	2%
	College grad	19.1	19.6	19.6	-1%	3%
Income	Less than $60k	18.1	19.1	19.2	1%	6%
	$60k-$100k	20.6	20.7	21.1	0%	2%
	$100k or more	19.8	20.0	19.8	-1%	0%
Tenure	Renter	18.4	19.9	20.5	4%	11%
	Owner	19.9	19.7	19.4	-3%	-2%
Region	Northeast	17.9	19.6	17.9	10%	0%
	South	21.0	21.4	21.3	2%	1%
	Midwest	18.6	18.0	19.3	-4%	4%
	West	19.1	19.0	19.7	0%	3%
All		**19.4**	**19.8**	**19.9**	**0%**	**3%**

We used the parameter estimates from the ECPP to predict the hours of care received by children in the CE samples to whom were imputed recipiency. The imputation quality can be evaluated by comparing the subgroup averages in the ECPP with their imputed counterparts in the CE Interview and Diary samples, respectively (Table 4-5 above). For the ECPP sample, recipients receive, on average, 19.3 hours of care per week. By construction, the imputed average in both the CE samples is practically identical to the ECPP average. None of the 29 subgroups in the table show a discrepancy larger than the 15 percent rule-of-thumb threshold we mentioned earlier. The deviations range from –4 percent (–5 percent for the Diary sample) to +10 percent. We conclude that the imputations are of a high quality in ascertaining the subgroup averages that may be of most interest to the research and policy community.

4.2 Care received by older adults

4.2.1 Background

Care received by older adults from people who do not live with them and without payment became of concern to economic policymakers in the mid-1980s. The interest was mainly because most such care was provided by women in their late–middle age helping their parents or parents-in-law with daily living tasks. Much more attention has been paid to the implications of this type of gendered-care arrangement in industrialized Western Europe than in the United States. Population aging is a more pressing problem in the former group of countries. Women should maintain or increase their labor force participation to reduce the growing dependency ratio. However, caregiving responsibilities toward elders can reduce women's labor force participation, especially when state support and assistance for older adults are being rolled back (see Moussa 2019 for a recent review).

Early studies on this topic in the United States examined the relationship between women's labor supply and parental care responsibilities using data on women who had such obligations, which prevented them from reliably estimating causal effects, as pointed out by Ettner (1995). In light of the then-existing policy concerns regarding the impact of societal aging, the SIPP collected information (including time spent) on caregiving for people who did not live in the respondent's household from *all* respondents in the mid-1980s using a topical module. The module was administered in Wave 7 of 1996, 2001, 2004, and 2008 SIPP panels and the data has been used for estimating the causal effect of parental caregiving on labor supply and formal employment on caregiving (e.g., He and McHenry 2016; Maestas, Messel, and Truskinovsky 2020). Unfortunately, the topical module on the care of nonhousehold members appears to have been discontinued in the SIPP's recent (i.e., post-2014) rounds.

A key disadvantage of the SIPP data is the lack of detailed information regarding the care receivers who live outside the home. The Health and Retirement Survey launched a module on functional limitations in the early 2000s, providing a breakthrough. Information was collected from the standpoint of the care receiver (respondent). Also, detailed information (e.g., sex, employment status, etc.) was collected about the caregivers related to the respondent. The joint information on care receivers and caregivers gave rise to extensive literature, including gender roles and the division of eldercare responsibilities among siblings (e.g., Grigoryeva 2017).

4.2.2 Data

4.2.2.1 Health and Retirement Survey (HRS) as the donor file

To our knowledge, the best source of information for our measurement purposes is the Health and Retirement Survey (HRS). The HRS uses a nationally representative panel sample of roughly 20,000 individuals over the age of 50. It is sponsored by the National Institute on Aging and the Social Security Administration and conducted by the University of Michigan. For an overview of the HRS, see Fisher and Ryan (2018). We identified the amount of time nonhousehold individuals spend caring for persons aged 50 years and over and for their spouses by using the 2016 HRS.

Some limitations of the HRS data for our exercise should be noted at the outset. First, it does not cover all adults but only older adults; therefore, we estimate a subset of adult care. The most sizeable omission here may be the nonmarket care received by nonelderly disabled persons. Second, the amount of time is reported by the care receiver rather than the caregiver and using the recall and not the diary method. Both aspects of data collection in the HRS may give rise to some (unknown) bias. Third, the questions on care received are not posed to all respondents but only to those who report some health difficulty or disability, thus rendering the care received by older people not afflicted by such problems impossible to measure. Finally, the scope of questions on the duration of care is limited to assistance with certain activities, as described below. As a result, even though a higher proportion of HRS respondents in 2016 report receiving help with cleaning or maintenance of their residence than with activities related to health difficulty (19 vs. 12 percent), no information is available on the hours of help with housework or whether the help was paid or unpaid.

The HRS module on time spent by helpers for the respondents begins by asking them a set of screening questions to identify if they face difficulties with everyday activities due to health or physical problems (e.g., walking several blocks, sitting for about two hours, etc.). Those that answered “yes” were asked about difficulty with Activities of Daily Living (ADLs) due to physical, mental, emotional, or memory problems.[45] All respondents were asked about difficulty with Instrumental Activities of Daily Living (IADLs).[46] Those who reported difficulty with ADLs or IADLs were asked whether they received any

[45] ADLs include dressing; walking across a room; bathing or showering; eating, including cutting up food; getting in and out of bed; and using the toilet.

[46] IADLs include preparing a hot meal, shopping for groceries, making phone calls, taking medications, and managing money.

help with the activities. A helper may be someone or an organization the respondent reported as assisting. Questions were asked about the helper's identity and the respondent's relationship to each helper (up to 7 helpers for ADL and 15 for IADL). From the information available in the HRS regarding the household members of the respondent, we can identify whether the helper in question resides with the person who received the help.[47] We can also ascertain whether the assistance is obtained from nonmarket sources (by excluding helpers related to the respondent as an "employee of a facility" or a "paid helper").[48]

Further questions were asked about the frequency and duration of help received from each nonmarket helper outside the respondent's household. We estimate the time spent for each respondent from the responses to the following questions. The survey offers the respondent three options to report the frequency of help received from each helper: the number of days during the last month, the number of days per week during the last month, or every day during the last month (respectively, QG070, QG071, and QG072). Once the respondent answers, they are asked about the hours per day they received help on the days they received help (QG073). We used the frequency of help and hours per day to calculate each helper's weekly hours of help to the respondent.[49] The total hours of nonmarket, nonhousehold help received by the respondent from all helpers were then obtained by summing the hours each helper gave to the respondent.

The HRS data show that the recipients of help are predominantly women (77 percent). A little less than half (44 percent) of the recipients are 81 years or older, and a majority (70 percent) are not currently married. The recipients also belong disproportionately to the bottom quintile of the income distribution (56 percent). The recipients are also disproportionately non-White and less educated compared with nonrecipients.

4.2.2.2 Comparing HRS and CE

A comparison of HRS and CE samples is not straightforward due to fundamental differences in design. The HRS follows individuals over 50 and their spouses, while the CE comprises a cross-section of consumer units. We first constructed a potential "HRS pool" from the Interview and Diary samples to

---

[47] If the helper is the spouse/partner of the respondent, we can identify whether they live together by utilizing the variable "2018 living arrangement status" (QLIVARR) in the so-called tracker file (Trk2018tr_r). If the helper is someone else, we can identify whether they live in the respondent's household by ascertaining whether the person is included among the household members of the respondent and a resident of the household. For this purpose, we use the file that contains the records of household members (H18pr_mc) and the variable "residency status updated" (QX056_MC).

[48] We use the variable in the helper-level file representing the relationship of the respondent to the helper (QG069) to perform this identification.

[49] For those who reported the frequency of help from a given helper as the number of days during the previous month, we multiplied the days and hours per day and divided the product by four to obtain weekly hours.

compare with the HRS. A person in a consumer unit was assigned to the HRS pool if they were a reference person over the age of 50, the spouse of a reference person over 50, or any other person in the unit over 50. It should be noted that, for couples in the last category, we cannot account for their relationship (e.g., characterize them as a married couple or identify the spouse's age) in our modeling.

Table 4-6 Demographic composition of the population over 50 years of age by survey

Characteristic	Subgroup	HRS 2016	CE - Interview	CE - Diary
Gender	Men	47	46	47
	Women	53	54	53
Age	51 to 60yrs	40	39	39
	61 to 70yrs	32	32	32
	71 to 80yrs	18	19	20
	81yrs or more	10	10	9
Household income quintile	Lowest	20	20	20
	Second	20	21	20
	Third	20	20	19
	Fourth	20	18	19
	Highest	20	22	22
Family type	Couple-only	46	42	43
	Single-person	25	22	22
	A couple and other adults	18	21	20
	Other households with 2+ adults	11	15	14
Race	Nonwhite	26	29	30
	White	74	71	70
Education	Less than high school	14	12	12
	High school	29	27	26
	Some college	26	28	28
	College	31	34	34
Tenure	Renter	30	21	23
	Owner	70	79	77
Division	New England	5	5	6
	Mid-Atlantic	12	15	14
	East North Central	16	13	14
	West North Central	8	5	5
	South Atlantic	22	21	21
	East South Central	6	7	6
	West South Central	9	11	11
	Mountain	8	7	7
	Pacific	14	16	15
Total		**100**	**100**	**100**

We compare the HRS 2016 sample with the CE samples (for 2019) in Table 4-6 in terms of their (weighted) composition. The samples are pretty much the same concerning gender, age, and distribution of household income across quintiles. However, some difference is found in the distribution of people across types of families. Compared to the CE samples, more people live in couple-only and single-person households as recorded in the HRS. We speculate that part of the problem here may be the difficulty of defining the relationships between the people in the HRS pool in the CE samples that were discussed before. Also, the HRS appears to have a lower proportion of White individuals and college graduates than the CE samples. The largest discrepancy is in home ownership. In the HRS, we observe 30 percent of people as renters while their shares are 21 and 23 percent, respectively, in the Interview and Diary samples. The difference may be driven by the fact the homeownership variable in the HRS is derived from the response to the question, "Do you [and your] [husband/wife/partner] own your home, rent it, or what?" Therefore, a father over 50 years of age who lives in a home owned by his son would not be recorded as a homeowner in the HRS, while that person would be coded as living in an owned home in the CE samples.[50] In sum, however, we think the differences in the sample composition are not sizeable enough to affect the imputation quality.

4.2.3 Models and results

According to the HRS data, the share of older adults who receive nonmarket care from people outside their household is only 3 percent. We reckon that 38 percent of individuals in both CE samples are in the HRS pool; hence, the number of people who would be affected by the imputation is a little over 1 percent of the population. This is similar to the situation we discussed for the preschoolers earlier in this chapter. These aspects of the problem—the small number of people affected, the pronounced subgroup differences in incidence and hours, and a limited role for intragroup variation—suggest that the imputation strategy we adopted for preschoolers would also be appropriate for older adults. We now outline the results of implementing that strategy.

[50] The HRS question allows responses "lives rent-free with relative/employer/friend" and "other" to the question regarding home ownership.

Table 4-7 Estimates of the binary logit model (dependent variable is the recipiency of care by persons over 50 years of age)

Variable	Estimate	Standard error
Intercept	-5.15**	0.200
Gender (reference group: female)		
Male	-0.84**	0.103
Age (reference group: 51 to 60 years)		
61 - 70 years	0.27**	0.130
71 - 80 years	1.04**	0.146
81 or more years	2.39**	0.134
Household type (reference group: composite household[a])		
One-person household	0.48**	0.090
Couple-only household	-0.31**	0.156
Race or ethnicity (reference group: non-Hispanic white)		
Nonwhite	0.65**	0.111
Education (reference group: College graduates)		
Less than high school	1.09**	0.156
High school	0.47**	0.168
Some college	0.40**	0.161
Home tenure (reference group: Owner)		
Renter	0.67**	0.098
Division (reference group: South Atlantic)		
New England	-0.30	0.252
Mid Atlantic	0.08	0.164
East North Central	0.24*	0.137
West North Central	0.14	0.141
East South Central	0.26	0.227
West South Central	0.44**	0.181
Mountain	-0.08	0.198
Pacific	0.12	0.151

Note: N = 18,927. The unit of observation is a person over 51 years of age. ** - significant at the 5 percent level, * - significant at the 10 percent level. Unmarked estimates are not statistically significant. Standard errors were calculated using the "stratum" and "cluster" variables that account for the complex survey design and are available in the data file. a – Composite household: households with three or more persons (two of which form a couple) and households with two or more people unrelated via marriage

Source: Authors' estimates from the public-use data file of the Health and Retirement Survey, 2016.

The recipiency model estimated from the HRS data is shown in Table 4-7. We used survey weights provided in the data file in our estimation. The independent variables used in the estimation are all dummy variables derived from the categorical variables shown above in Table 4-6. Our estimates indicate men are less likely to receive care than women. Further, the probability of receiving care increases with age, compared to people in the 51 to 60 age group. Men may be less likely than women to receive care, because women are overrepresented in the older age groups due to their longer life expectancy. Compared to older adults in composite households with at least two adults, people who live by themselves are more likely to receive care, while couple-only families are less likely to receive care. This is intuitive because a couple may be able to meet their care needs independently and are likely to be younger. College graduates appear to be less likely than people of lower educational attainment to receive care; this may reflect a cohort effect as the younger groups are more likely to be college graduates than older groups. Lower wealth also seems to be associated with a higher probability of receiving care, as indicated by the positive sign of the coefficient associated with renters. It is worth noting that all the variables, except some pertaining to geographical location, display high statistical significance. However, as in the case of care for preschoolers, we should recall that we are not factoring in proximity to family or friends, preferences, and other unobserved factors that can influence the recipiency of nonmarket care from people outside the household.

The next step in the imputation procedure involved assigning recipiency to persons over 50 years of age in the CE samples. We utilized the coefficients of the binary logit model shown in Table 4-7 above to accomplish the goal. A stochastic component was added to the prediction by assigning recipiency based on a draw from a Bernoulli distribution, using the predicted score for each individual as a parameter for the draw. We assessed the resulting imputation's quality by comparing the CE samples' imputed care incidence with the HRS' observed incidence (Table 4-8).

Table 4-8 Observed and imputed recipiency of care (percent) among persons over 50 years of age

Characteristic	Subgroup	HRS	CE - Diary	CE - Interview	Deviation from the HRS (%)	
					CE - Diary	**CE - Interview**
Gender	Men	1.6	1.4	1.6	-9%	5%
	Women	4.6	4.2	3.9	-10%	-15%
Age	51 to 60 years	1.3	1.4	1.3	5%	-1%
	61 to 70 years	1.7	1.5	1.5	-8%	-10%
	71 to 80 years	3.7	3.3	3.7	-13%	0%
	81 years or more	15.1	12.8	12.0	-15%	-21%
Household type	Couple-only	1.4	1.4	1.5	0%	10%
	One-person	6.5	5.9	6.0	-9%	-9%
	Other	3.2	2.8	2.5	-12%	-22%
Race	Nonwhite	5.0	4.1	4.5	-19%	-11%
	White	2.6	2.4	2.2	-8%	-13%
Education	Less than high school	8.6	7.0	7.2	-18%	-16%
	High school	3.4	3.5	3.2	4%	-6%
	Some college	2.5	2.2	2.4	-13%	-2%
	College	1.3	1.5	1.5	20%	20%
Tenure	Renter	5.8	5.5	5.6	-6%	-3%
	Owner	2.1	2.1	2.1	1%	3%
Division	New England	2.1	2.2	1.9	8%	-7%
	Mid-Atlantic	3.3	2.7	3.0	-18%	-9%
	East North Central	3.1	3.1	2.7	-1%	-13%
	West North Central	3.0	2.8	2.3	-6%	-25%
	South Atlantic	2.7	2.3	2.7	-17%	-2%
	East South Central	3.2	2.9	3.3	-9%	2%
	West South Central	5.3	4.8	4.4	-9%	-17%
	Mountain	2.2	2.1	1.7	-5%	-26%
	Pacific	3.5	2.9	2.9	-17%	-18%
All		**3.2**	**2.9**	**2.9**	**-10%**	**-10%**

Note: HRS refers to the Health and Retirement Survey, 2016.

Overall, we estimated that the recipiency rate in the HRS was 3.2 percent. The imputed recipiency rate in the Diary and Interview samples was slightly lower at 2.9 percent. The table also shows the observed and imputed recipiency rates for 26 population subgroups. Once again, using the rule-of-thumb of 15 percent deviation, we find that six and seven subgroups exceeded the threshold in the Diary and Interview samples, respectively. In both samples, three "problem" cases were associated with the Census division of the household's location. Two educational attainment groups also show divergence in both samples: our procedure seems to assign higher-than-observed incidence to college graduates and the converse for people who have not completed high school. In the Diary sample, the remaining problem case was an understatement of recipiency among non-Whites. In contrast, the imputation in the Interview sample was much closer to the observed value. However, the imputation for “Other” types of households and the oldest age group in the Interview sample fell short of their observed counterparts in the HRS.

While the 15 percent divergence threshold we are using is instructive, it is important to note for the present application that it may exaggerate divergences because of the relatively low base values. Consider, for example, the most considerable deviation of 26 percent for the Mountain division in the Interview sample. The observed incidence in that division is only 2.2 percent. In comparison, our imputation assigns an incidence of 1.7 percent—a discrepancy of 0.5 percentage points that may not be considered as far off the mark. In sum, we judge that imputations are reasonably accurate in reproducing the subgroup differences in the recipiency of nonmarket care by older adults from people outside their household.

Once the recipients were identified in the manner described above, we proceeded to impute the hours of care received by individuals in the CE samples. Just as for preschoolers, we found that a Poisson model produced the best predictions. The independent variables used in the regression were dummies reflecting the same characteristics used in the incidence model, although the reference group was changed for some. Survey weights were used in the estimation. The results are shown in Table 4-9.

Table 4-9 Estimates of the Poisson model (dependent variable is the weekly hours of care received by a person over 50 years of age)

Variable	Estimate	Standard error
Intercept	3.49***	0.219
Gender (reference group: female)		
Male	-0.26*	0.190
Age (reference group: 51 to 60 years)		
61 - 70 years	-0.53**	0.267
71 - 80 years	-1.01***	0.261
81 or more years	-0.44***	0.165
Household type (reference group: single-person household)		
HH with 2+persons	0.26	0.210
Couple-only HH	0.08	0.237
Race or ethnicity (reference group: non-Hispanic white)		
Nonwhite	0.47**	0.181
Education (reference group: Less than high school)		
High school	-0.07	0.214
Some college	-0.12	0.240
College	-0.24	0.375
Home tenure (reference group: Renter)		
Owner	-0.11	0.169
Division (reference group: South Atlantic)		
New England	-0.66	0.558
Mid Atlantic	-0.24	0.239
East North Central	-0.17	0.265
West North Central	-0.60*	0.362
East South Central	-0.28	0.293
West South Central	-0.17	0.281
Mountain	-0.71***	0.292
Pacific	-0.38	0.272

Note: N = 877. The unit of observation is a person over years of age. ** - significant at the 5 percent level, *** - significant at the 1 percent level. Unmarked estimates are not statistically significant. Standard errors were calculated using the cluster and strata variables that account for the complex survey design and are available in the data file.

Source: Authors' estimates from the public-use file from the Health and Retirement Survey (2016).

Estimates for some covariates in the hours model have an opposite sign compared to the same covariate in the incidence model. We found a similar reversal of signs for preschoolers too. Perhaps the most striking is age. The incidence model showed a positive association between age and the probability of receiving external care. However, the hours model shows the opposite: conditional on the other covariates, the average hours of care received tends to fall with age. There are also sign reversals for some of the geographical divisions. Except for the coefficient on couple-only households, all parameter estimates are significant at the 5-percent level.

The final step in our imputation procedure was assigning weekly hours of care to those identified earlier as recipients in the CE samples. We used the coefficients of the Poisson model described above to attain this goal. As usual, we examined the imputation quality by comparing the observed and imputed average hours (Table 4-10).

Table 4-10 Observed and imputed weekly hours of care received by adults 50 years and older

Characteristic	Subgroup	HRS	CE - Diary	CE - Interview	Deviation from the HRS (%) CE - Diary	CE - Interview
Gender	Men	16.3	18.7	18.3	15%	12%
	Women	24.0	23.6	24.3	-2%	1%
Age	51 to 60 years	18.9	19.7	20.9	4%	10%
	61 to 70 years	13.9	11.6	13.0	-17%	-6%
	71 to 80 years	21.6	20.3	20.3	-6%	-6%
	81 years or more	27.0	29.5	29.2	10%	8%
Household type	Couple-only	20.6	22.8	23.3	11%	13%
	One-person	20.2	19.3	20.2	-4%	0%
	Other	26.9	26.5	26.2	-1%	-3%
Race	Nonwhite	28.4	28.9	28.4	2%	0%
	White	18.1	17.7	18.2	-2%	1%
Education	Less than high school	26.2	25.8	28.4	-1%	8%
	High school	21.0	23.5	23.6	12%	12%
	Some college	19.4	18.3	19.4	-6%	0%
	College	18.1	19.9	16.7	10%	-8%
Tenure	Renter	23.9	23.5	24.1	-1%	1%
	Owner	20.3	21.7	21.9	7%	8%
Division	New England	15.1	17.7	14.5	18%	-4%
	Mid-Atlantic	22.6	23.4	24.0	4%	6%
	East North Central	21.3	19.2	21.4	-10%	0%
	West North Central	14.5	14.7	14.2	1%	-2%
	South Atlantic	29.3	27.1	30.1	-8%	3%
	East South Central	21.6	17.4	19.3	-20%	-10%
	West South Central	25.3	27.5	24.1	9%	-5%
	Mountain	14.1	11.7	13.1	-17%	-7%
	Pacific	21.8	23.8	19.8	9%	-9%
All		**22.6**	**22.5**	**22.8**	**-1%**	**1%**

According to our estimates from the HRS, people who reported receiving nonmarket care from persons outside their household received, on average, 22.6 hours of care per week. Our imputation, by design, hits the mark at 22.5 and 22.8 hours per week in the Diary and Interview samples, respectively. Turning to the 26 population subgroups depicted in the table, we observe that none of the imputations for the Interview sample fall outside the 15-percent rule-of-thumb threshold of deviation. For the Diary sample, there are four instances where the deviations crossed the threshold, of which three were found in the Census division. The remaining one occurred for the age group 61–70. None of the four cases showed a deviation exceeding 20 percent. As a result, we can conclude that the imputations succeed in replicating the overall and subgroup averages of care received.[51]

4.3 Summary

We estimate that, of a population of nearly 21 million children under six years of age in 2019, about 1.9 million (9.3 percent) receive, without any payment, regular care from relatives or nonrelatives who do not live with them. For those who receive such care, the average hours of care received amount to 19.4 hours per week. In the aggregate, this translates to approximately 38 million hours per week. Or, in terms of 35-hour-per-week jobs, a little over one million.

The care received by the children is vital to them. Additionally, through their caring activity, the kith and kin (especially grandmothers, who constitute the majority of such caregivers) can also potentially facilitate the children's parents' pursuit of education or employment, thus enabling families to sustain or enhance their standard of living. The benefits of having well-cared-for children and promoting productive activity of their parents extend beyond the families of children to the broader community and society.

Turning to similar care received by older adults, about 3.6 million persons over 50 receive care from people (relatives or nonrelatives) outside their household for no payment. The incidence of such nonmarket care was only 3.2 percent in the total population of 112 million older adults in 2019. We estimate that the average hours of care received by the recipients is about 22 hours per week. Therefore, about 80 million weekly hours of caring activity occur in this sphere, which would amount to nearly 2.3 million full-time (35 hours per week) jobs. The HRS data shows that daughters or daughters-in-law carry out most of this caring activity. Their "hidden" labor is part of the social costs of caring for older adults. Just as with the care of preschoolers, the beneficiaries of the care of older adults are also not the immediate recipients of care alone.

The scope of our present study is to incorporate the contribution made by these types of nonmarket care, usually left out of the analysis, into a comprehensive measurement of nonmarket care. From this

[51] Our imputations were not completely effective in terms of replicating the averages by household income quintiles. A reason could be the differences in the definition of household income. In the HRS, household income includes only the income of the respondent and their spouse while in the CE samples, incomes of all members are included. About 70 percent of people over 50 years of age live in couple-only or single-person households (see Table 6). For these households, the income definitions in the two surveys coincide. However, that leaves out 30 percent of people for whom there may be discrepancies across surveys. In spite of these differences, we found that the average hours imputed for the bottom 40 percent and the top 20 percent of households in the income distribution in the two CE samples were identical to the observed averages in the HRS. For the third quintile, we understate the average hours and for the fourth quintile, we do the opposite. An obvious choice for those interested in looking at income-based differences in the hours may be to lump the third and fourth quintiles in the CE samples into a single category.

perspective, the estimates of nonhousehold, nonmarket care we provide underestimate the actual volume of such activity. We do not account for the care that older children (especially those in the 6 to 13 years age group) may receive from people outside their household—their relatives, neighbors, or nonrelatives. It is reasonable to think that regular care received by the children in this age group will primarily take place for a limited period before the start or end of the school day. On average, the duration of care received by older children is thus likely to be shorter than that of younger children. Yet, the population of children in the 6 to 13 age group is almost double that of the younger group. Consequently, we may be missing a considerable number of hours.[52] To get an approximate idea of how much we are missing, we can compare the ECPP aggregate hours of care to the total weekly hours of care given by respondents in the ATUS to nonhousehold children.[53] The latter is 127 million hours, about 3.3 times higher than its ECPP counterpart. While the ECPP collects only information on regular care received by preschoolers, the ATUS measure encompasses all children under 18 years and does not distinguish between regular and irregular care.

Finally, we are not accounting for the nonhousehold, nonmarket care received by adults in the 18 to 50 age group. Systematic data collection on the nonmarket help received by those with disabilities who live in households is probably required to fill this gap. The ATUS collects data on time spent caring for nonhousehold adults, but it does not contain any information regarding the age of the adults. Therefore, we cannot develop separate estimates for the care received by adults in the 18 to 50 age group, just as we cannot estimate the care received by nonhousehold children in the 6 to 13 age group. We also noted that our source of data for the care of older adults, HRS, covered only help received in daily activities (such as bathing or paying bills) by those with some mental or physical difficulties. The gap in coverage leaves an unknown but perhaps nontrivial amount of the care received by older adults from our measurement exercise. We may use the ATUS benchmark of aggregate hours of care given to have a rough idea of the omission's magnitude. Estimates from the 2019 round of ATUS indicate that help given by respondents in the ATUS to nonhousehold adults amounted to 119 million weekly hours.[54] This amount is about 1.5 times higher than that recorded in the HRS.

[52] While the ATUS collects information on care given by respondents to nonhousehold children, it does not record the age of nonhousehold children.

[53] Our assumption is that, in principle, the aggregate hours of care given and received should be identical. We calculated the aggregate hours of care given as those reported under the category of "Caring and helping nonhousehold children" delineated in the ATUS tables published by the BLS.

[54] We estimated this using the category of "Caring and helping nonhousehold adults" in the ATUS tables published by the BLS.

5 Consumption Expenditures and Household Production

5.1 Introduction

The purpose of this chapter is to present a preliminary analysis of the imputations. Our estimates in this chapter were calculated from the synthetic file created by statistically matching the third-quarter CE Interview file with the ATUS, augmented by the imputations from the ECPP and HRS for nonmarket help received from people outside the consumer unit. The CE sample for the third quarter has a little over 5,000 consumer units. All the calculations were done using the weight variable "finlwt21." The matching procedure is described in Chapter 3, and the imputations from ECPP and HRS are discussed in Chapter 4.

We first indicate the extent to which the inclusion of household production alters the average value of expenditures. In light of the elaboration of the differences between our preferred measure and the "BEA" measure (Chapter 2), we examine the impact using both measures. Our preliminary analysis is also extended to assess subgroup disparities. Here, we present evidence of the much higher contribution made by women to household production. Further, we also consider how much the inclusion of household production affects the measured differentials among households differentiated by two key demographic characteristics—size and composition of the household as well as by the race and ethnicity of the reference person. Next, we turn to the distributional profile of household production, i.e., the gap in the average value of household production across the deciles of expenditures. To better understand the distributional impact of adding the value of household production to expenditures, we also conduct decomposition analyses of the Gini coefficient by groups (with and without children) and by source (expenditures and categories of household production). Apart from analyzing overall distributional issues, synthetic data may also be used for analyzing subsets of spending and household production. We briefly illustrate this point by considering food and childcare. The final section offers some concluding comments and caveats.

5.2 The overall impact of the inclusion of household production

We begin with an overview of the average monetary values of various categories of household production, namely, unpaid (nonmarket) help received from persons not residing in the household and home production by household members for their household. The former is further split into help received in caring for young children and help received in caring for older adults (Table 5-1). We report two sets of monetary values.[55] Those labeled "Preferred" are based on our household production definition and our valuation method. Alternatively, "BEA" indicates results derived from using what we described earlier as the "BEA method." This approach excludes supervisory care and uses a single "generalist" wage, instead of specialist wages as we do, to convert time into dollars. We discussed the differences between the two approaches and wage rates earlier in Chapter 2.

The definition of expenditures shown in Table 5-1, line 1 was derived by subtracting, from total quarterly expenditures,[56] certain expenditures on owner-occupied primary homes (mortgage interest, property taxes, maintenance, repairs, insurance, and other expenditures) and expenditures classified as miscellaneous, cash contributions, and personal insurance and pensions. We divided the resulting quarterly amount by three to get a monthly value and added the monthly rental value of owned primary home to obtain the estimates shown in line 1.

Specifically, we first calculated the sum of expenditures in the current quarter (TOTEXPCQ) and previous quarter (TOTEXPPQ) and divided it by three to obtain monthly amounts. In the next step, we subtracted certain categories of expenditure from the sum obtained in the second step. The categories are: mortgage interest, property taxes, maintenance, repairs, insurance, and other expenditures on owner-occupied primary home; expenses classified as miscellaneous, cash contributions, and personal insurance and pensions. Each expenditure item was formed by summing the expenditures in current and previous quarter and dividing the sum by three to obtain monthly amounts.[57] Finally, we added the rental equivalent for owned homes (RENTEQVX), which is available in the file as a monthly amount.

The expenditures we subtracted from total expenditures are also excluded by Garner et al. (2023). Similarly, our addition of imputed rent for own-homes also mirror their approach.[58] However, our measure does not reflect some key modifications made by Garner et al. (2023) to arrive at their measure

---

[55] We report monthly values of household production rather than weekly values because they match expenditures reported on a monthly basis. The monthly values were calculated by multiplying weekly values by 4.345. However, this implies the assumption that the weekly values we constructed are typical for all the weeks of the month. (See Section 3.4.)

[56] Total expenditures consist of expenditures in 14 major categories: food, alcohol, housing, clothing, transportation, healthcare, entertainment, personal care, reading, education, tobacco, miscellaneous, cash contributions, and personal insurance and pensions.

[57] The variables used to create the quarterly expenditures are as follows: Mortgage interest: MRTINTPQ+MRTINTCQ; property taxes: PROPTXPQ+PROPTXCQ; maintenance, repairs, insurance, and other expenditures on owner-occupied primary home: MRPINSPQ + MRPINSCQ; miscellaneous: MISCPQ + MISCCQ; cash contributions: CASHCOPQ + CASHCOCQ; and, personal insurance and pensions: PERINSPQ + PERINSCQ. Each expenditure item was divided by three to obtain monthly amounts.

[58] We did not make corresponding adjustments in constructing a consumption measure in our validation exercise using the PSID data because the categories we subtracted could not be readily identified and an estimate of imputed rent was not available.

of consumption. First, we do not remove net outlays and financing charges associated with vehicles because an estimate of service flow from vehicles is not available to replace them in the public-use CE files. Second, we did not net out education and medical expenditures in accordance with the ILO and OECD guidelines. Finally, we do not include imputed monetary values for in-kind transfers because the requisite imputed values are not available in the public-use CE files. Analyzing the impact of including household production on alternative measures of consumption would be interesting but it falls outside the scope of our current research.

Table 5-1 Average monthly values of expenditures and household production per consumer unit, 2019 (third quarter)

		Preferred*			BEA**		
Line	**Categories**	**No children**	**With children**	**All**	**No children**	**With children**	**All**
1	Expenditures***	4,402	5,490	4,726	4,402	5,490	4,726
2	Household production	1,889	5,433	2,989	1,836	3,564	2,368
3	Unpaid help from nonhousehold people	48	99	64			
4	Help with childcare for children under 6		79	24			
5	Help with adult care for adults over 50	48	20	40			
6	Home production of household members	1,840	5,334	2,925	1,836	3,564	2,368
7	Food preparation and clean-up	408	696	494	410	697	496
8	Housework (except food preparation)	1,377	1,648	1,458	1,377	1,648	1,458
9	Childcare		2,920	914		1,157	362
10	Active care		1,449	453		1,157	362
11	Supervisory care		1,471	461			
12	Not joint with home production		1,014	318			
13	Joint with home production		458	144			
14	Adult care	55	70	60	49	63	53
	Addendum:						
15	Augmented expenditures (Line 1 *plus* Line 2)	6,242	10,923	7,715	6,238	9,055	7,095
16	Augmented expenditures (as a percent of Expenditures)	142%	199%	163%	142%	165%	150%

Notes: *Includes unpaid help from nonhousehold people and supervisory childcare in the definition of home production. A specialist-wage method is used for valuation. **Excludes unpaid help from nonhousehold people and supervisory childcare in the definition of home production. A generalist-wage method is used for valuation. *** Excludes mortgage interest, property taxes, maintenance, repairs, insurance, and other expenditures on owner-occupied primary home; also excluded are expenses classified as miscellaneous, cash contributions, and personal insurance and pensions. Includes estimated monthly rental value of owned primary home.

In assessing the results, it is essential to differentiate between households with and without children (defined as persons under 18 years). In our estimates, most households have zero values for time spent on childcare. However, a sizeable minority (approximately 30 percent of consumer units) have positive childcare time values. On average, time spent on household production differs drastically between the two groups of households. The difference is manifested in the results we now discuss.

The effect of adding the value of home production to expenditures is shown in line 15 with the label "Augmented expenditures." Our method increases the average expenditures of the entire sample by 63 percent, while the increase under the BEA method is smaller at 50 percent.[59] In either scenario, the impact of home production on measured average consumption expenditures is enormous.

Notably, both methods show that the impact on the average value of units with children is higher than that for all units: the increase is virtually 100 percent with our approach and 65 percent with the BEA method. On the other hand, for units without children, both methods produce approximately the same percentage increase of a little over 40 percent. Since standard expenditures (shown in line 1) are relatively higher for households with children than without, another way to interpret the finding is that accounting for home production widens the gap further in favor of the former group. Our method does this more so than the BEA approach.

The fact that preferred and BEA methods produce the same result for those without children indicates that our inclusion of nonmarket care received from outside the household had an insignificant impact. Indeed, the estimates shown in line 3 show that the average amount was pretty small relative to the rest of home production.[60] Further, the use of specialist instead of generalist wage had a negligible effect on the home production of household members in consumer units without children, as shown by the practical concurrence of average values under both valuation strategies (line 6). The trivial effect is not surprising because, under our method, the generalist wage is used for valuing housework (other than cooking) and

[59] The estimated value of "Nonmarket services" (mostly household production) in the BEA's satellite account amounted to 22 percent of personal consumption expenditures in 2020 (Bridgman et al. 2022, Table 2). The lower share in NIPA as compared to our measure is not surprising because of the broader definition of expenditures used in the national accounts.

[60] As noted in Chapter 4, our estimate of nonmarket help from outside the household is downward biased because we did not include care received by older children and adults under 50 with disabilities or serious illness.

shopping. Another reason is that the specialist wage for cooking is not much different from the generalist wage, as described in Chapter 2.

We should now address why our estimate of the average value of home production for households with children is notably higher (52 percent) than the BEA counterpart, as shown in line 2. Here, just as for households without children, we find virtually the same average values for cooking and other housework for the same reason. The driving force behind the discrepancy between the two methods is childcare. Two factors determine the difference in the average value of childcare. The first and less important reason is that the specialist wage for active childcare (used by us) is somewhat higher than the generalist wage (employed in the BEA method). We can observe the impact of this difference in the valuation of active childcare (line 10), which shows that our estimate is about 25 percent higher than the BEA counterpart.

But, the difference between our estimate of childcare and the BEA estimate shown in line 9 is a much larger 152 percent! Thus, the second and more important reason for the discrepancy is the inclusion of supervisory care in our definition and its omission in the BEA definition of childcare. The value of supervisory care is almost as large as active childcare even though it is valued at one-half the rate of active childcare because the time allocated to supervisory care compensates for the differential in the hourly rate. About 70 percent of the value of supervisory care stems from care that is not overlapping with other household production and the remaining from care that overlaps with other household production. Defining what counts as childcare trumps how it is valued (at least under the current wage structure) in determining the impact of including home production in a broader measure of consumption.[61]

5.3 Subgroup disparities

5.3.1 Gender

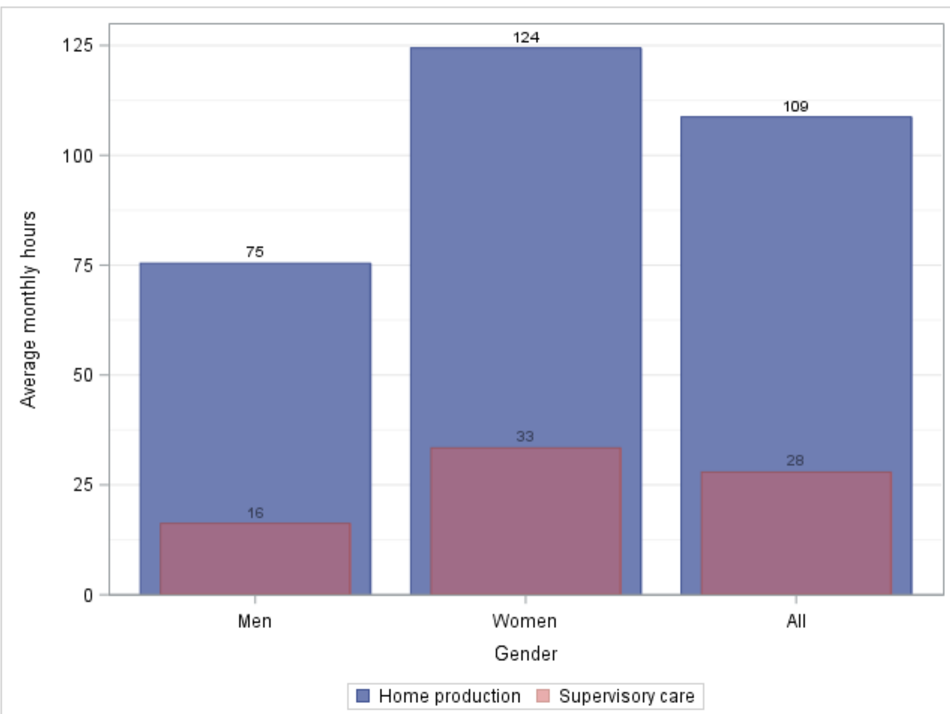
We first focus on the well-known gap between men and women regarding the time spent on household production. Our estimates show that including supervisory care widens the gap further (Figure 5-1). While women work about 124 hours per month on home production, men work about 40 percent less, or 75 hours. But, the gap in supervisory childcare is even higher: Men's engagement in this type of caregiving is 52 percent less than women's (16 vs. 33 hours per month). The gap in our context implies that, by working longer hours providing household services, women make a greater contribution to enhancing the consumption of their households. As a result, we estimate that women account for the

[61] The definition of time spent on supervisory care was given in Section 2.2 (see especially Table 2-1 and text below the table). We discuss the valuation of supervisory care in Section 2.3 (please see especially the paragraph above Table 2-2 and text below the table).

overwhelming share—78 percent—of the total $386 billion added to monthly consumption expenditures in 2019.

Interestingly, women's contribution to aggregate hours of home production is the same (about 78 percent of the 28.6 billion monthly hours). Thus, the larger contribution of women in augmenting consumption is due to their sheer volume of work that far exceeds that of men. Gender differences in the composition of home production (e.g., women spending more time than men on childcare, valued at a higher rate) or differences among the specialist wages we used in valuation played practically no role in shaping this outcome.

Figure 5-1 Average monthly hours of home production and supervisory care by persons 15 years and over by gender, 2019



Note: N=10,357. Calculations are based on the individual-level synthetic file for the Interview sample (2019: Q3) and imputations via statistical matching. All men and women are included in the calculation including those who did not engage in any home production.

5.3.2 Size and composition of consumer units

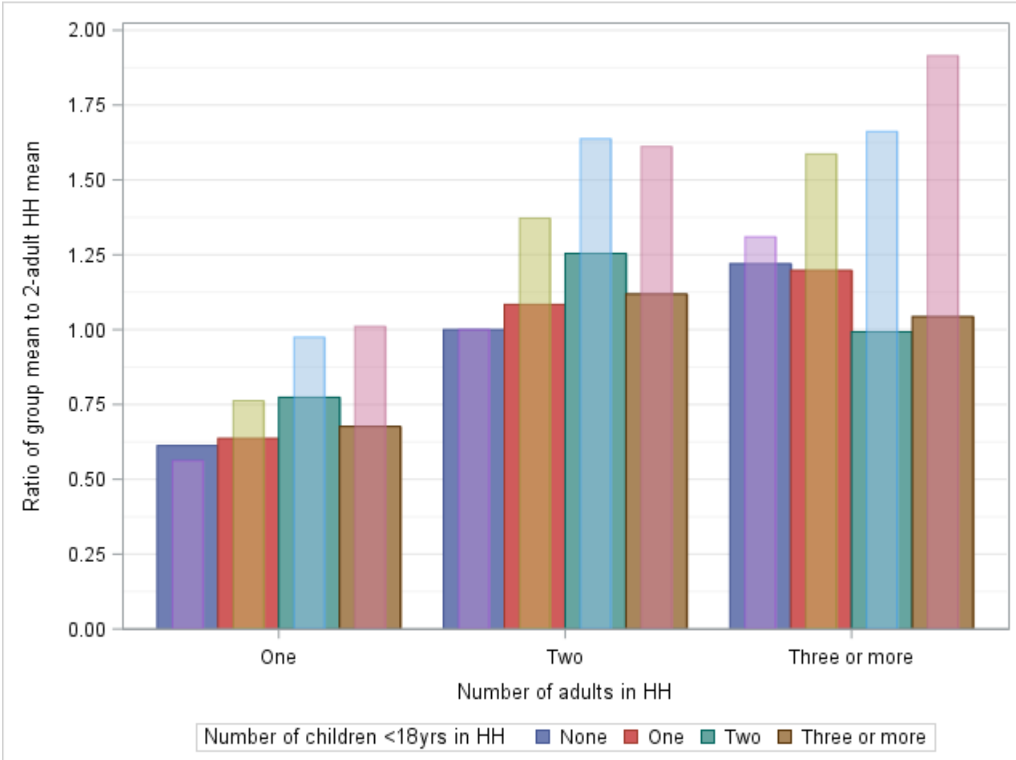
We have already considered how the incorporation of household production exacerbates the measured disparity in consumption between households with and without children in favor of the former. The presence of children, in general, creates more demand for home production, and such demand would increase with the number of children, other things remaining constant. Older children can, of course, be providers of home production too. On the other hand, the amount of household production also tends to rise, all else remaining the same, with the number of adults in the household because of its effect on both demand and supply of home production. As is well-known, the supply side is affected positively not just because more people can engage in home production; more people also bring in the possibility of someone in the household dedicating themselves to taking care of the house and dependents. We now present some evidence on the extent of these variations among consumer units differentiated by the number of adults and children. We are interested in how much the variations in home production alter the pattern of differences in market consumption expenditures between consumer units. It is a separate matter to assess these differences normatively by placing them against differences in "needs" with the latter, codified by an explicit or implicit equivalence scale. The question of equivalence scales will have to be confronted for building a measure of poverty that incorporates deprivation in home production. Properly treating this issue falls outside the scope of our present research.[62]

In Figure 5-2, we have shown the differences in the average value of market expenditures and augmented expenditures among consumer units differentiated by the number of adults and children. We have adopted the group with the largest share in the total number of consumer units, the consumer unit with two adults and no children, as the "base group";[63] the average value of each group is expressed as a ratio to the average value of the base group. Hence, by construction, the average value for the two-adult-no-child household equals one. The ratios calculated using standard expenditures are represented by the broader bars, while the ratios corresponding to augmented expenditures are shown by the thinner bars.

[62] One of us has argued strongly against using either the Betson three-parameter scale or square-root scale in the context of home production elsewhere and proposed an alternative equivalence scale (Folbre, Murray-Close, and Suh 2018). A different approach is taken in Zacharias et al. (2019, Appendix B) as a part of constructing a measure of time and consumption poverty. Here, a reference group of households is identified first and thresholds for household production, differentiated by the number of young children, older children, nonelderly adults, and elderly adults are estimated next for the reference group by means of a nonlinear regression model. The estimated time thresholds are then applied to all households for ascertaining their time-poverty status. This approach is similar in spirit to the standard procedure of constructing absolute income or consumption-poverty thresholds.

[63] The two-adult–no-child households made up 31 percent of all consumer units. A close second position was taken by the single-person households (30 percent).

Figure 5-2 Relative average expenditures by the number of adults and children in the household, standard vs. augmented expenditure definition, 2019



Note: Average expenditures are expressed as a ratio to the average expenditures of households with two adults and no children. We estimate that for the latter, the average values for standard and augmented expenditures were, respectively, $5,106 and $7,446. The broader bars show the ratios calculated with standard expenditures, and the thinner bars indicate the ratios associated with augmented expenditures (standard expenditures plus our preferred value of household production. The latter includes supervisory childcare and nonmarket help from outside the household, and valuation is performed according to specialist wages.

We find that the ordering of household types changes notably. Relative expenditures are highest for the two-adult–two-child households when home production is ignored (1.25). But, with the inclusion of home production, the highest average is for the group with the largest number of adults and children (1.91). Without home production, the group's average was very close to that of the base group (1.04). A less spectacular increase is found for three-plus–adults–two-child households, whose relative consumption widens from parity of standard expenditures to 1.66 with augmented expenditures as the yardstick. The lowest average value occurs for the same type—single-person household—with standard and augmented expenditures. But, their distance from the base group widens further when home production is included (0.61 vs. 0.56).

As we would expect, given the number of adults, the gradient with respect to the number of children becomes clearly positive with the inclusion of home production. This change is perhaps most visible for households with three or more adults: average expenditures are generally *lower* in consumer units with more children, but the opposite holds when augmented expenditures are considered. For households with two adults and households with a single adult, three or more children are associated with lower average expenditures relative to two-children households. However, no such decline is discernible with augmented expenditures as the yardstick.[64]

5.3.3 Race and ethnicity

As we have just seen, the presence of children in the household—as well as the number of adults and children—has a decisive impact on the amount of home production. We now examine briefly how these factors shape racial differences in augmented consumption expenditures. We use the standard approach of assigning groupings for households based on the group of the household reference person. We designate Hispanics of all races as "Hispanic." Non-Hispanic Whites and Blacks are referred to hereafter as "White" and "Black." Non-Hispanic householders of races other than White or Black (primarily Asian) are grouped as "Other." Because of sample size limitation and our use of imputed variables, we may not obtain reliable estimates if we were split the Others into Asians and everyone else, though that would have been desirable. Our tabulations showed that Whites, Blacks, Hispanics, and Others comprised 66.7 percent, 13.3 percent, 13.6 percent, and 6.4 percent of consumer units in 2019, respectively.

[64] It might be useful to note that these estimates have implications for the ratio of adult consumption to child consumption: household consumption can increase while adult consumption declines.

Table 5-2 Racial differences in monthly average values of household production and their potential determinants

Race/Ethnicity	Value of home production	Hours of home production	Implicit unit value of home production	Share of households with children	Number of children	Number of adults
	A. All consumer units					
White	2,754	205	13.4	24%	0.44	1.82
Black	2,690	200	13.5	34%	0.69	1.78
Hispanic	4,013	292	13.7	50%	1.04	2.18
Other	3,801	270	14.1	41%	0.71	2.11
	B. Consumer units with children					
White	5,391	399	13.5		1.87	2.14
Black	4,835	354	13.7		2.04	1.97
Hispanic	5,783	418	13.8		2.09	2.37
Other	5,875	412	14.3		1.71	2.31

Note: Household production includes supervisory childcare and nonmarket help received from outside the household, and valuation is performed according to specialist wages. The implicit unit value is obtained by dividing the value of household production by the household production hours. Consumer units are assigned to racial groups based on the racial group of the household reference person. See the text above the table for the definition of the groups.

The average monthly value of home production was the highest for Hispanic consumer units (Table 5-2 Panel A, above). The Hispanic value was 46 percent higher than the average value for White households. An almost similar proportionate difference between the two groups exists in the average hours of home production. This suggests that the difference in the volume, rather than the implicit unit value of home production,[65] is the main driver behind the intergroup gap. Proximate factors related to the higher average hours for Hispanic consumer units are presented in the last three columns of the table. Half of all Hispanic households have children compared to only a quarter of White households. The average number of children is about 2.4 times more in Hispanic than White households (1.04 vs. 0.44), and the average number of adults is also larger (2.18 vs. 1.82), although the gap here is smaller, as we would expect.

Turning to the other groups, we observe almost no difference between Blacks and Whites regarding the average monthly value of household production. However, the share of households with children is higher for Blacks than Whites (34 percent vs. 24 percent), and the average number of children is also greater for Blacks (0.69 vs. 0.44). These factors should push Blacks' average hours of home production above that of

[65] We refer to the unit value as "implicit" because different components of home production are assigned different hourly values in our valuation procedure. Thus, substantial differences in implicit unit values would indicate corresponding differences in the composition of home production (e.g., a higher share of childcare in the total hours of home production for a group compared to other groups).

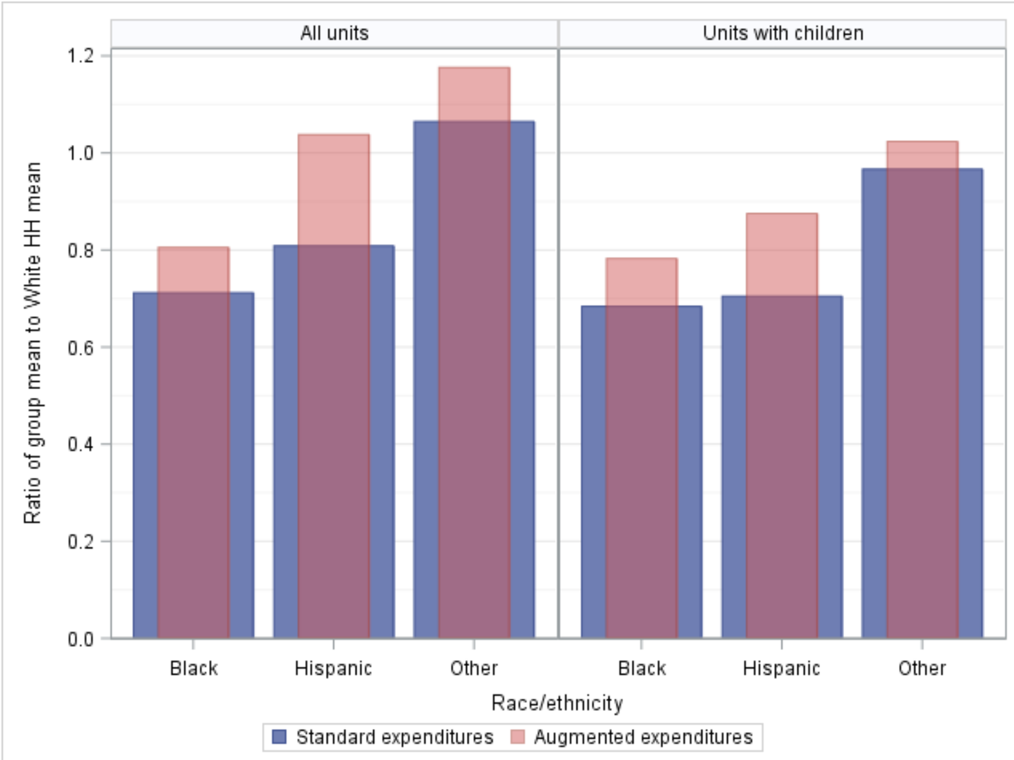
Whites. But, the average number of adults is lowest among Black consumer units (1.78), and this may help explain, at least partially, why their average hours are virtually the same as Whites'.

The same factor (i.e., the difference in the number of adults) also probably contributed to the gap between Blacks and Others. Though the average number of children is similar for the two groups (0.69 and 0.71), the number of adults is notably higher for Others (2.11 vs 1.78). The higher number of adults and the greater prevalence of households with children compared with Blacks (41 vs. 34 percent) help us understand why, despite having a similar average number of children, the average hours of Others are closer to those of Hispanics than Blacks. However, the average value of home production between Hispanics and Others shows a smaller gap than average hours because the implicit unit value of household production is highest for Others among the groups.

We find that racial differences in the monthly average hours of home production are notably less when we consider only consumer units with children (Table 5-2 Panel B, above). Clearly, the principal factor here is the drastically narrow range of differentials in the average number of children. Blacks continue to register the lowest average amount of time spent on household production, probably because they have the lowest average number of adults at 1.97, below the 2-adult benchmark. Perhaps as a consequence, White consumer units with 2.14 adults per household are found to have higher average monthly hours even though their average number of children is lower than that of Blacks (1.87 vs. 2.04).

The implications of incorporating home production for the picture of racial disparities in expenditures must also be considered. Given the decisive influence of the presence of children, we examined the differences among all consumer units and, separately, for consumer units with children. We use White households as the base group so that the average expenditures of consumer units in other racial groups are expressed as a ratio to the White household average.

Figure 5-3 Relative average expenditures by race/ethnicity of the household, standard vs. augmented expenditure definition, 2019



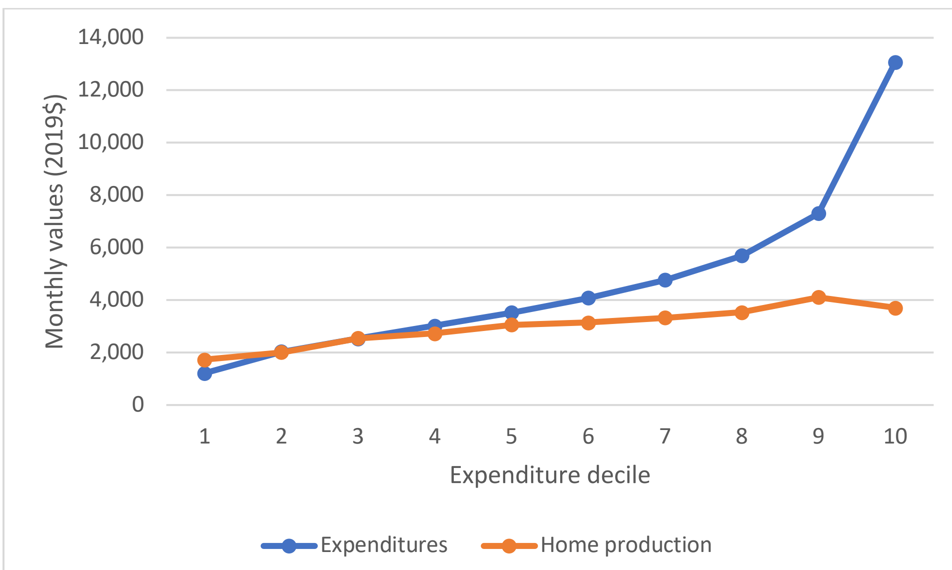
Note: Average expenditures are expressed as a ratio to the average expenditures of households with a White household reference person. The average for White households without children was equal to $5,036 and $7,789 for standard and augmented expenditures, respectively. White families with children registered mean standard expenditures of $6,253 and mean augmented expenditures of $11,644. The broader bars show the ratios calculated with standard expenditures, and the thinner bars indicate ratios associated with augmented expenditures (standard expenditures plus our preferred value of household production). The latter includes supervisory childcare and nonmarket help received from outside the household, and valuation is performed according to specialist wages.

Perhaps the most surprising finding is the ranking reversal between Whites and Hispanics when we examine the whole sample. In terms of standard expenditures, the average Hispanic value is about 20 percent lower than Whites; however, Hispanics are marginally better off than Whites in terms of augmented expenditures. As we saw before, the greater prevalence of consumer units with children, the higher average number of children, and the higher average number of adults among Hispanics were the proximate factors behind the rather large wedge in the average value of household production between Hispanics and Whites. We now see that the lead is sufficient to close the gap in measured expenditures between the two groups. On the other hand, when we consider the subsample of units with children, White households maintain their lead over Hispanics even after accounting for household production. But, the size of the gap is lower. The reason behind this result is the smaller White–Hispanic gap in household production for the subsample compared with the whole sample, which, in turn, can be attributed to the lower gaps in the average number of adults and especially in the average number of children.

5.4 Distribution of household production

Not surprisingly, we find that the value of household production (our preferred measure) is distributed much more equally than expenditures across the deciles[66] of expenditure (the expenditure definition used in calculating the deciles is "standard expenditures," as defined in Table 5-1 above). This is illustrated in Figure 5-4, which plots the average spending and average value of household production across expenditure deciles. Interestingly, average expenditures and home production values are similar for the bottom half of the expenditure distribution. Beyond the 50th percentile, home production rises slightly, and expenditures increase much faster up to the 90th percentile. Finally, at the top, there is a sharp increase in spending but a slight decline in home production.

Figure 5-4 Average monthly values of expenditures and home production by expenditure decile, 2019

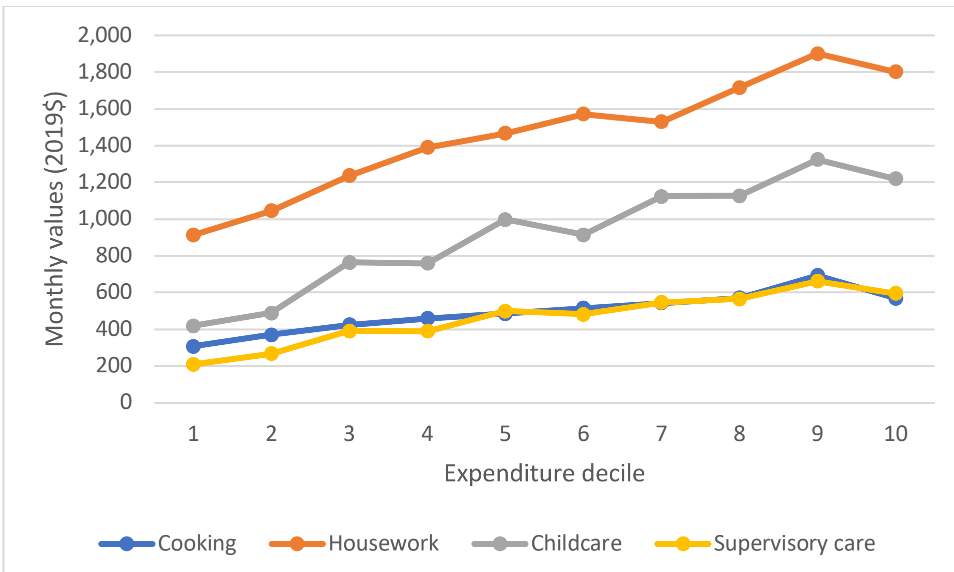


The distributional profiles of the major components of home production, cooking, housework (including shopping), and childcare also suggest much lower inequality than the distribution of expenditures (Figure

[66] We are not using an equivalence scale in defining the deciles (see below, Section 5.3.2, for a discussion of equivalence scales). The deciles are constructed such that each consists of approximately the same weighted number of consumer units.

5-5). Throughout the distribution, housework is the dominant component, followed by childcare. A subset of childcare is supervisory care, and its average value across deciles is quite close to that of cooking, especially in the top half of the distribution. Childcare and its subset, supervisory care, show the sharpest gradient. Using the simple ratio of the top decile's mean to the bottom decile's mean, we find the value to be 2.1 for home production as a whole but 2.9 for childcare. In comparison, the same ratio is equal to 10.7 for expenditures.

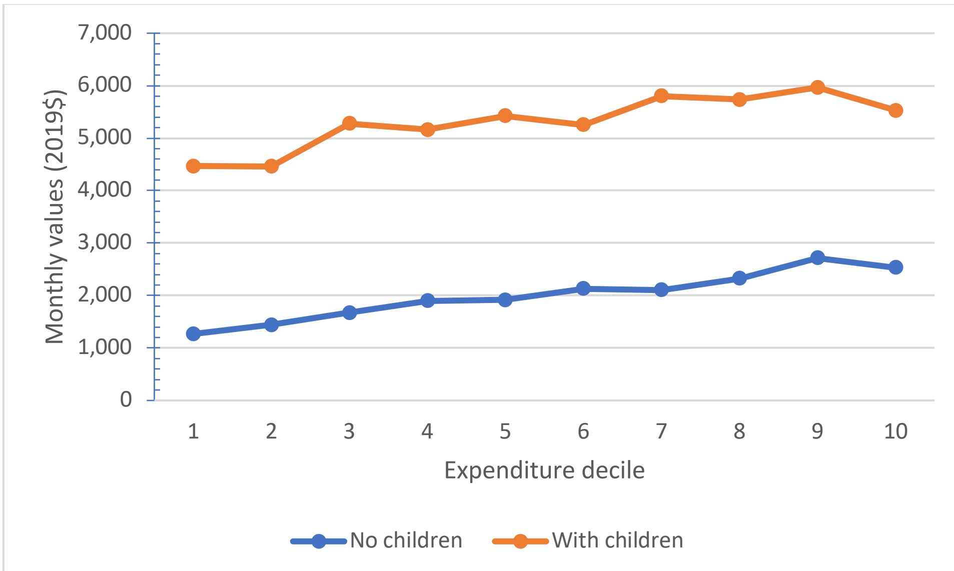
Figure 5-5 Average monthly values of the major components of home production by expenditure decile, 2019



Note: Childcare includes supervisory care.

Families with children are overrepresented in the higher deciles. The compositional effect contributes to the positive correlation between the rank in the expenditure distribution and the value of household production because the latter tends to be higher for families with children than those without. We can see the composition effect in how, as a proportion of the number of consumer units without children, families with children change across the distribution. The value is only 0.17 in the bottom decile, but that value rises to almost 0.5 by the middle decile and 0.65 in the top decile. Hence, it is crucial to examine the distribution of the value of home production within each group (Figure 5-6). Our estimates show that the distribution of home production is actually more equal among families with children across the expenditure deciles than among households without children. While there are notable bumps in the average value for families with children at the third and seventh deciles, the curve connecting the values is notably flatter than its counterpart among households without children. Using our crude metric of the ratio of the top decile's mean to the bottom decile's average, we find a value of 2.0 for consumer units without children, practically identical to the value of 2.1 we found for the whole sample. However, the corresponding value for families with children is only 1.24.

Figure 5-6 Average monthly value of home production by expenditure decile and presence of children (under 18) in the household, 2019



Note: Expenditure deciles are defined using the cutoffs for the whole sample and not for subgroups.

The large gap in the mean values of household production between those with and without children indicates that adding home production to expenditures will result in a wider gap between the two groups in the augmented measure compared to the standard estimate of expenditures. Intergroup inequality will therefore increase when we incorporate household production. However, because the distribution of household production is far more equal than the distribution of spending, the inequality in the augmented measure within each group will be lower than in the standard measure. Given the larger values of home production and the greater equality in its distribution across expenditure deciles among households with children, we expect their decline in inequality to be much sharper compared to households without children. We also expect the reduction in within-group inequality to overwhelm the increase in between-group inequality and, therefore, the measured inequality to decrease with the introduction of home production.

Table 5-3 Subgroup decomposition of Gini coefficients for expenditures and augmented expenditures by the presence of children in the household and measure of home production, 2019 (Gini points)

Line		No children	With children	All	Within-group	Between-group	Overlap
1	Expenditures	36.0	33.9	35.7	20.0	5.5	10.3
	Preferred						
2	Augmented expenditures	31.8	23.8	32.5	15.9	13.1	3.5
3	Change (Line 1 *minus* Line 2)	4.2	10.1	3.2	4.1	-7.6	6.8
	BEA						
4	Augmented expenditures	31.9	25.6	31.4	16.8	9.0	5.6
5	Change (Line 1 *minus* Line 4)	4.1	8.3	4.3	3.2	-3.5	4.7

The estimates confirm our expectations (Table 5-3). For the entire sample and our preferred measure, we find that the Gini coefficient falls by 3.2 percentage points, from 35.7 to 32.5, when we add home production to the standard definition of expenditures. We also find that the Gini coefficient declined by 10.1 points for households with children and by a lower amount of 4.1 points for households without children. Analogous behavior is also observed for the alternative BEA measure: Inequality in the augmented measure is lower, and the decline in inequality is more sizeable for families with children.

However, a comparison of lines 3 and 5 shows that for the sample as a whole (shown in column "All"), the decline in the Gini coefficient is greater for the BEA than for our definition. A decomposition of the Gini by subgroup sheds insight into the reason behind the difference. Specifically, the increase in between-group inequality when home production is included turns out to be substantially bigger under our measure (a rise of 7.6 vs. 3.5 Gini points).[67] The greater increase in within-group inequality is enough to offset the slightly bigger declines in within-group inequality and overlapping inequality for our measure relative to the BEA. In turn, the higher jump in between-group inequality is driven by the inclusion of supervisory care, which is as large as active childcare in monetary value, in our measure and its omission in the BEA definition.

[67] The decomposition can be expressed as: $G = G_b + G_w + O$, where G is the Gini coefficient, G_b is between-group inequality, G_w is within-group inequality and O is the overlap or residual term. G_b is simply the Gini coefficient of group averages while $G_w = \sum_{i=1}^{n} G_i\, s_i p_i$, i.e. a weighted average of the within-group Gini coefficients with weights given by the product of the share of group i in the total quantity of the variable for which we are measuring inequality, s_i, and the share of group i in the total population (number of households), p_i (see, e.g., Yao 1999).

Table 5-4 Decomposition of the Gini coefficient by components of consumption (Gini points) by measure and presence of children, 2019

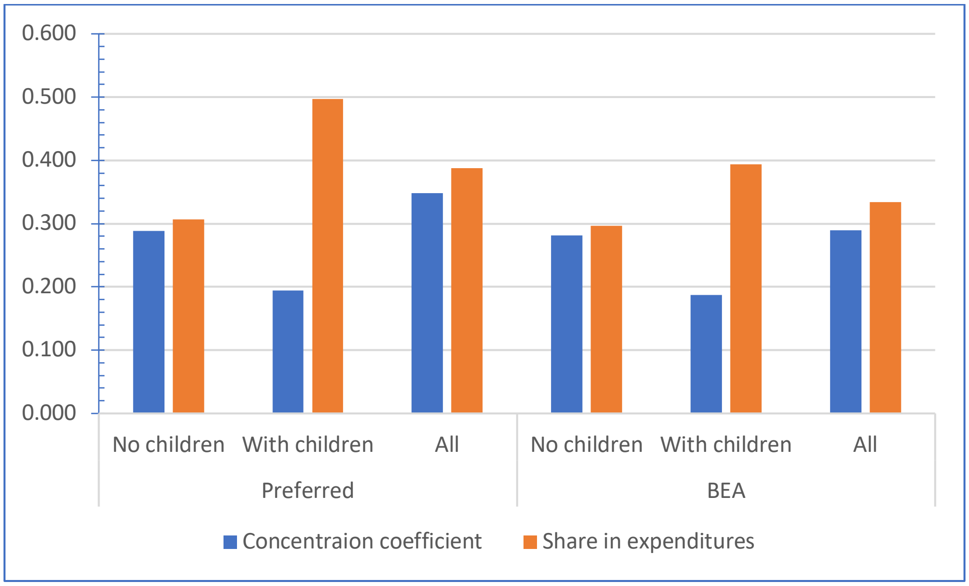
		Preferred			BEA		
Line		No children	With children	All	No children	With children	All
1	Expenditures	23.0	14.1	19.0	23.5	18.2	21.7
2	Home production	8.9	9.6	13.5	8.4	7.4	9.7
3	Augmented expenditures	31.8	23.8	32.5	31.9	25.6	31.4
	Addendum:						
4	Cooking	1.6	1.0	1.6	1.6	1.1	1.7
5	Housework	6.6	2.9	4.7	6.4	3.7	5.4
6	Active childcare		3.1	3.4		2.3	2.3
7	Supervisory care		2.1	3.2			
8	Other	0.6	0.5	0.6	0.4	0.2	0.3

Note: The numbers in lines 1 and 2 represent the contribution (in Gini or percentage points) of expenditures and home production to the Gini coefficient of augmented expenditures shown in line 3. The numbers in lines 4 through 8 show the contribution of each listed category (e.g. housework) to the Gini coefficient shown in line 3.

A decomposition of Gini by source can help us better understand the relative roles of the different components of home production in driving the reduction in measured inequality (Table 5-4).[68] We can observe by reading down the numbers in the "No children" columns (preferred and BEA) that the contribution of household production and its components to inequality in augmented expenditures is roughly the same for both definitions. As we would expect, the difference between the measures stems from how distribution changes within the group of families with children and their impact on the overall distribution. Within this group, our measure's expanded definition of childcare contributes to the larger inequality-reducing effect of home production. Simultaneously, it loses some of its inequality-reducing strength because our measure enhances the between-group component of inequality (Table 5-3). As we saw before, the inequality in BEA augmented expenditures is lower than in preferred augmented expenditures. Our estimates in Table 5-4 capture this effect in the higher contribution of home production to overall inequality (line 2, columns labeled "All") in our measure than in the BEA's (13.5 vs. 9.7 Gini points). We also show further, in lines 6 and 7, the greater contribution of total and supervisory childcare to inequality under our definition.

[68] We use the standard decomposition by source formula: $G = \sum_{i=1}^{n} A_i = \sum_{i=1}^{n} c_i\, s_i$, where G is the Gini coefficient of expenditures, A_i is the contribution of component i to the Gini, c_i is the concentration coefficient of component i with respect to expenditures, and s_i is the share of component i in expenditures (Yao 1999).

Figure 5-7 Factors affecting the contribution of home production to consumption inequality by alternative measures, 2019



The numbers shown in Figure 5-7 help interpret the reasons behind the higher contribution of home production to inequality with an expanded definition of childcare. As we know, the contribution of a component of expenditures to the Gini of total expenditures (standard or augmented) is simply the product of its concentration coefficient with respect to total expenditures and the component's share of total expenditures. A comparison of "With children" bars in the two definitions clearly shows that home production's larger contribution to within-group inequality of preferred measure is almost entirely due to its higher share in augmented expenditures (because of the inclusion of supervisory care). The concentration coefficients are practically identical for both measures. However, a comparison of the "All" bars demonstrates that the concentration coefficient of the preferred measure is also notably higher (0.35 vs. 0.29). In turn, the higher concentration coefficient reflects the greater inequality in the distribution of home production across the expenditure distribution because of the larger gaps between families with and without children.

5.5 Applications to food and childcare

We hope the synthetic data will also be useful for investigating issues that require information regarding the joint distribution of specific categories of household production and expenditures. As noted earlier (see Appendix A of Chapter 3), the favorable conclusion about our imputation strategy that emerged from our validation exercise was restricted to the joint distribution of total household production and total expenditures. In this section, we examine specific subsets of spending and household production to illustrate the synthetic data's potential uses rather than offer a full-fledged study of the particular topics.

We begin with food and assess the existence of an inverse relationship between the share of food expenditures in total food expenditures (a measure of the degree to which food requirements are met via outsourcing) and household hours of "cooking."[69] Next, we explore the relationship between childcare expenditures and hours of childcare. For married-couple families, we are focused on the intrahousehold division of childcare responsibilities and investigating whether our synthetic data indicate that dual-earner families who spend, on average, more on childcare than "male-breadwinner" families also display a lower share in wives' total family hours of childcare. We also examine if families with children and a single employed reference person who purchase childcare services engage in a lower number of hours of childcare than their counterparts who do not. Our findings are generally consistent with economic intuition. However, they must be considered tentative because we do not account for the data's special features, and some of our estimates are generated from a relatively small number of observations, as discussed below.

5.5.1 Food

The average household expenditures on food constituted about 16 percent of the monthly average total expenses of about $4,700. Spending on food consumed at home accounted for two-thirds of total food expenditures, and spending on food away from home made up the remainder.[70] Households also perform a considerable amount of household production devoted to the preparation of food and drinks as well as cleaning up afterward (referred to as "cooking" here).[71] On average, cooking accounted for about 17 percent of monthly total household production hours of 220. As expected, it is also an almost universal activity with a participation rate of 94 percent.

---

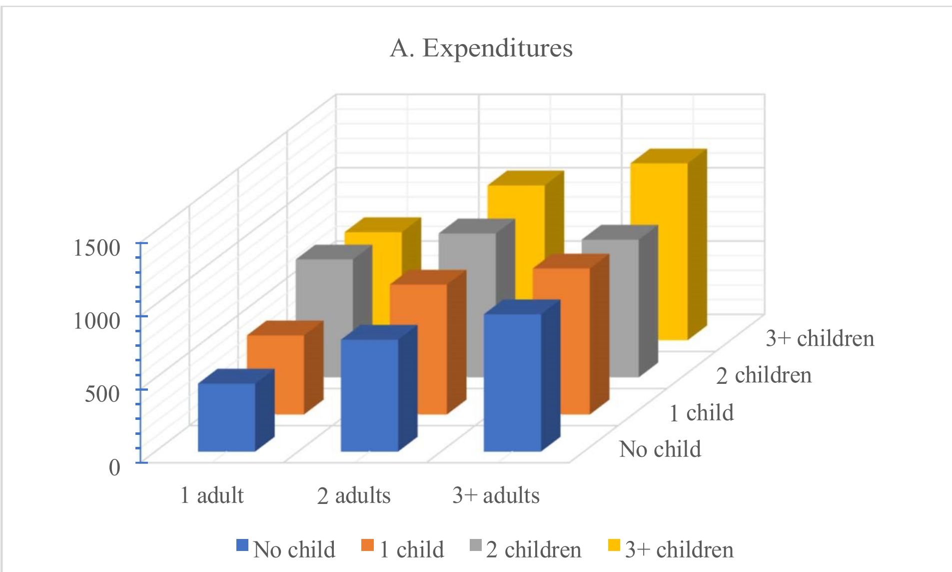
[69] Our question is related to, yet different from the mainstream approach of viewing "eating" as a commodity produced by food expenditures and "time" inputs, which includes, in addition to the hours of food preparation etc., also the time spent on eating and drinking (see, Hamermesh 2007 for an early discussion and Gardes 2019 for a recent reformulation). The main difference is that we do not consider the time spent on eating and drinking as contributing to a product. Our interest is also not to analyze the reasons why people choose outsourcing of food over cooking at home but to look at the effect of the mix (outsourcing and own preparation) of food expenditures on hours of cooking.

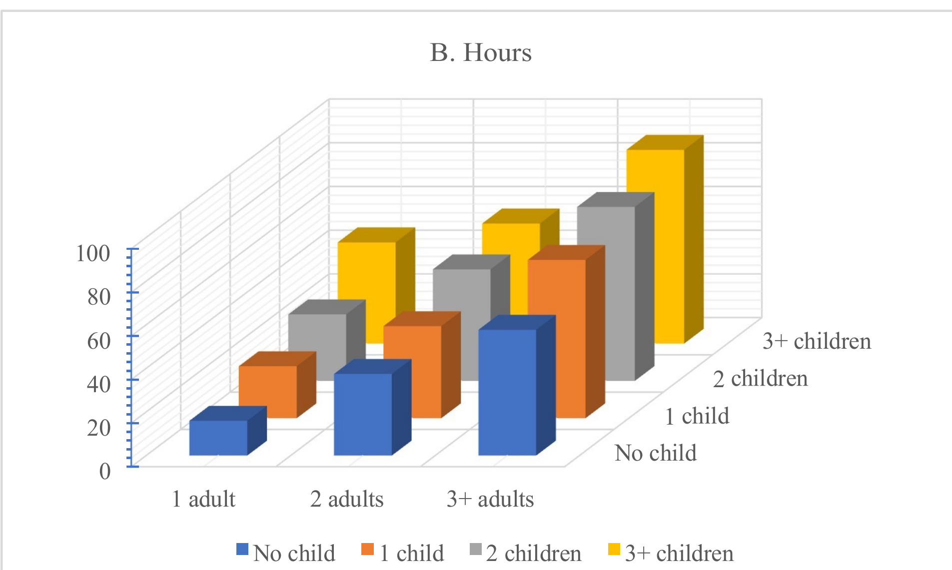
[70] The food expenditure variables are taken from the public-use FMLI file, FMLI193 (the variable names are shown in the parentheses): food expenditure is the sum of expenditures on food at home (FDHOME) and food away from home (FDAWAY).

[71] Please see Table 2-1 for the ATUS categories included in our measure of "cooking" time. Basically, it consists of hours of food and drink preparation; food and drink presentation; and kitchen and food clean-up.

Food expenditures differ considerably by household size and composition (Figure 5-8, Panel A). Several factors are at play, and let us note a few. Variations in need exist among households because more people require more food, with children requiring, generally, less food than adults. Then, there are economies of scale in expenditures (e.g., the price per quart of soft drink purchased in a 20oz vs. 2-liter bottle), which can dampen the variation. On the other hand, the means to afford more and higher-priced food varies across households, partly as a function of the number of earners, which also exerts an influence over the variation in food expenditures.

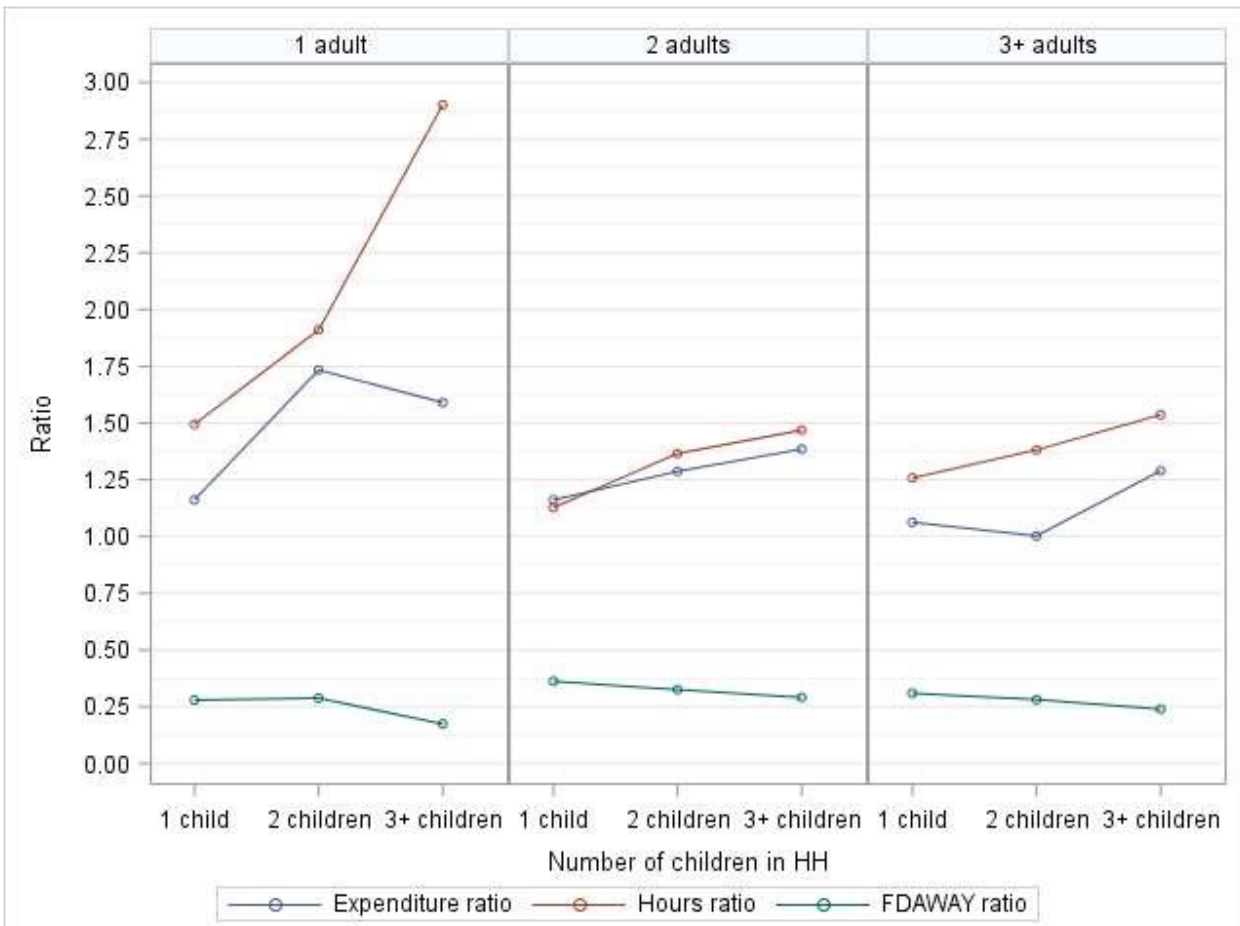
Figure 5-8 Average monthly food expenditures and hours of cooking by number of adults and children in the household, 2019





Variations in the hours of cooking are also affected by the size and composition of the household (Figure 5-8, Panel B). Some of the proximate factors at work here are the same as those discussed earlier in Section 5.3.2. Admittedly, economies of scale exist in cooking. But, this may be somewhat offset by differences in preferences among household members. The presence of children and their number in the household alters the quantity and mix of food, which can influence the cooking hours. Both expenditures and hours of cooking generally display a positive gradient with respect to the number of adults and children. However, an additional factor that can influence the gradient of hours and expenditures is the degree to which food requirements are met via outsourcing, i.e., food away from home. An illustration of this effect may be seen by comparing the gradients of hours and expenditures, both with respect to the number of children and holding the number of adults in the household constant (Figure 5-9).

Figure 5-9 Average monthly food expenditures, average monthly household hours of cooking, and share of expenditures on food away from home in total food expenditures by size and composition of households, 2019



Note: "Expenditure ratio" refers to the ratio of each group's average expenditures to the subgroup's average expenditures with no children. Similarly, the "Hours ratio" is calculated by dividing the mean household hours by the mean hours of the subgroup with no children. "FDAWAY ratio" is the ratio (share) of average expenditures on food away from home in total food expenditures.

The red line in each panel of the figure represents, for a given number of adults, the average food expenditures of households with a certain number of children as a ratio of the average spending of households with no children. We can observe along the blue line the same information but for hours of cooking. For example, in families with one adult and one child, the average food expenditure is 1.16 times higher, and the average hours of cooking are 1.49 times higher than that of the corresponding averages of one-adult households with no children. The spending and hours gradients appear to coincide roughly for the two-adult households. But for families with one adult and families with three or more adults, the hours' gradient lies markedly above the expenditure gradient, i.e., average hours increase relatively more than average expenditures when the number of children increases by one. A potential explanation would lie in the extent to which food supply is outsourced, i.e., a greater proportion of food requirements are met via expenditures on food away from home. This proportion is consistently higher for families with two adults than households with one adult and households with three or more adults. For example, in households with two children, the share of expenditures on food away from home in total food expenditures in one-adult, two-adults, and three-or-more-adult families constituted 29, 33, and 28 percent, respectively.

Table 5-5 OLS estimates of a simple model of household hours of cooking

Variable	Parameter estimate
Intercept	17.82**
	(0.90)
One child	7.60**
	(1.29)
Two children	15.20**
	(1.35)
Three or more children	21.57**
	(1.70)
Two adults	20.46**

	(0.94)
Three or more adults	44.80**
	(1.28)
Share of food away from home	-5.69*
	(1.91)

Note: The dependent variable is monthly household hours of cooking, N = 5,335. ** indicates significance at the 1% level, and * indicates significance at the 5% level. No adjustment has been made to the standard errors to account for the imputed nature of the dependent variable. Neither were the replicate weights in the CE file applied.

A crude test of the proposition that the household hours of cooking were negatively correlated to the share of expenditures on food away from home in total food expenditures was carried out utilizing a simple OLS regression with dummies for the number of children (taking households with no children as the base group) and the number of adults (households with one adult serving as the base group) serving as the control variables (

Table 5-5). The coefficient on the share of food away from home suggests that a percentage point increase will likely lead to a decline of six hours per month (roughly 11 minutes per day) in time devoted to cooking. However, it should be noted that we have not accounted for the fact that the independent variable is imputed. We also have not applied the replicate weights required to estimate robust standard errors in the CE data. Appropriate adjustments to reflect these properties of the data may affect the statistical significance of the parameter estimates reported here.

5.5.2 Childcare

Assessing the impact of purchasing childcare services on the hours of childcare performed by parents or other members of the children's household is rather difficult. A primary reason is that the US does not currently have a nationally representative survey that contains information on expenditures and hours of market and nonmarket childcare. Our imputed datafile allows some tentative assessments regarding this question because it includes information on childcare expenditures[72] collected in the CE and imputed hours of nonmarket care received by the household's children. However, we do not have information on the hours of childcare purchased. Hence, we cannot identify the potential hours that could be "saved" by those who buy the services.

[72] We used the "BBYDAY" variable in the FMLI public-use file to calculate expenditures.

Further caveats arise because of the imputation itself. These are especially relevant for childcare. The potential buyers of childcare services—consumer units with children (defined as persons under 18 years)—constitute a minority, though a substantial one (30 percent of the sample). The actual buyers make up 20 percent of the units with children or 6 percent of all consumer units. We should note that the quality of imputation for such a small segment is not likely to be as good as for larger subgroups and for categories of household production with much larger participation rates (e.g., housework or cooking). To avoid small cell sizes, we use all five quarters relevant to 2019.[73]

With these caveats in mind, we examine the relationships between employment, the hours of childcare provided by the reference person or spouse of the reference person, and the purchase of childcare services. We focus on employed households—households in which the reference person, spouse, or both are employed—because the expenditure data shows that almost all childcare expenditures are incurred by employed households. We use the standard family typology: heterosexual married couples, households with a single female reference person, and families with a single male reference person. For the first type of family, our main interest is in the intrahousehold division of childcare responsibilities (as reflected in the total hours of childcare provided by all family members). For single-person families, we assess the impact of spending on childcare services on household childcare hours.

We consider married couples first, restricting our sample to families with exactly two adults (the reference person and spouse) to avoid complications arising from the contributions made by other relatives or nonrelatives in the household to childcare. Our interest is in examining how average childcare expenditures and hours of childcare vary between families with an employed and nonemployed wife. The strongly gendered division of childcare, coupled with the fact that almost all of the husbands in our sample are employed full-time, justifies the focus (e.g., Foster and Kreisler 2012). We estimate from the expenditure data that the proportion of families with an employed wife purchasing childcare services is approximately twice as much as those with a nonemployed wife (Table 5-6). Among families who report childcare expenditures, this category of spending makes up roughly 10 percent of total consumption expenditures for those with an employed wife compared to 5 percent when the wife is nonemployed.

[73] That is, we use "191x, 192, 193, 194, and 201" matched files.

Table 5-6 Employed married-couple families with children and two adults: Average monthly childcare expenditures and average monthly hours of childcare provided by the family, 2019

Characteristic	All	Wife not employed	Wife employed	Difference
Percent reporting childcare expenditures	26	16	29	13
Average expenditures among spenders	676	407	722	315
Percent of total expenditures	9.0	5.0	9.7	5
Household hours of childcare				
Total	295	350	278	-72
Supervisory	197	228	187	-42
Direct	99	122	92	-30
Wives' hours of childcare				
Total	179	237	161	-76
Supervisory	116	150	105	-44
Direct	63	87	56	-32
Husbands' hours of childcare				
Total	109	104	111	6.3
Supervisory	76	71	77	5.5
Direct	34	33	34	0.8

Notes: (i) The column "Difference" shows the result of subtracting values listed under "Wife employed" from "Wife not employed." (ii) The total household hours are greater (by a small amount) than the sum of the hours of husbands and wives because it also includes childcare rendered by family members (probably older children) between 15 and 17 years. (iii) Components may not add up to their respective totals due to rounding. (iv) There were 4,396 families, of which 1086 and 3310 had nonemployed and employed wives, respectively. (v) 1,120 consumer units reported spending on childcare, of which 170 and 950 had, respectively, employed and nonemployed wives.

Turning to hours of childcare, we observe that the average monthly hours are 72 hours less for families with an employed wife than a nonemployed wife, reflecting, to some extent, the higher spending on childcare services among the former. Both supervisory and direct care hours are lower. Divergent patterns of change are evident for wives and husbands. Employed wives' average hours of childcare are 76 hours less than their nonemployed counterparts. The intrahousehold division of childcare responsibilities shifts markedly in wives' favor with employment—the share of employed wives in total household hours of childcare is 58 percent, ten percentage points lower than that of nonemployed wives' average share. However, on average, husbands with an employed wife engage in childcare for about 6 hours more than those with a nonemployed wife. Most of the difference between them is accounted for by the higher average hours of supervisory care for husbands in dual-earner families. Our additional tabulation (not reported in the table) showed that 96 percent of the difference in supervisory care among husbands could be attributed to supervision that does not overlap with other household production (e.g., keeping an eye on the children while watching TV). Supervisory care that coincides with other household work (e.g., watching children while cooking) played a trivial role in shaping the difference.

Families with children and a single, employed reference person face different constraints. Most are households with one adult, thus eliminating the possibility of sharing childcare or household financial responsibilities with another adult. Since the constraints imposed by childcare are generally more severe for families with children under 13 (because they admittedly require supervision), we consider this group here. The caveats we outlined earlier regarding relatively small cell sizes apply naturally, with even greater force for families with a single reference person. Our estimates indicate that childcare expenditures are associated with fewer hours of childcare (Table 5-7). The reduction appears true for both single females and single males, amounting to roughly 40 minutes per day for females and 1 hour and 10 minutes for men. However, the cell sizes are pretty low for those who purchase childcare services (144 and 29, respectively). Single-female families purchase childcare services at about the same rate as dual-earner families (29 percent), while the rate among single-male families is lower at 21 percent (these estimates are not shown in the table).

Table 5-7 Families with children under 13 years and a single, employed reference person: average monthly childcare expenditures and average monthly hours of childcare provided by the family, 2019

	Family hours of childcare		Average expenditures of families purchasing childcare	
Gender of the reference person	**Does not buy childcare services**	**Buys childcare services**	**Amount**	**Percent of total expenditures**
Female	193	174	401	10
Male	156	120	498	8

Notes: (i) Family hours of childcare refer to the sum of direct and supervisory care hours provided by all family members 15 years or older. (ii) 615 families are included in the sample for the estimates shown in the table, of which 475 and 140 have female and male reference persons, respectively. Buyers of childcare services in the sample were 173, of which 144 and 29, respectively, had female and male reference persons.

5.6 Summary

If a measure of consumption is to include the value of household production, the average value of that measure will be substantially higher than the value before the inclusion. We found this to be true irrespective of which measure—ours or the BEA's—is used. However, the extent of the increase in expenditures is much larger when our measure is employed (1.5 vs. 2.0 times). The main reason is that we include supervisory childcare while the BEA measure excludes it. Defining childcare is far more crucial than how it is valued for determining the size of household production. In contrast, the inclusion of unpaid help from non-household persons seems to have a minor impact on the average value of household production. As noted before, this may be partly due to the inadequacies in the available data to capture the full extent of such help.

The additional consumption that is made available to household members via household production is not made possible through household members' equal contribution to domestic labor. Indeed, this reflects the well-known gender disparity in time spent on household activities and childcare. Our estimates showed that women contributed 78 percent of the extra potential consumption due to household production.

We briefly examined how measured disparities between population subgroups are affected by the inclusion of household production. The number of adults and children strongly influences the volume of household production. Without including home production, we find that the average expenditures were highest for households with two adults and two children. When household production is added to expenditures, the highest average value is observed for the group with three or more adults and children. It also became apparent that the presence of children caused the re-ranking. For households without children, the ranking is unaffected by the inclusion of household production. Our scrutiny of the change in racial disparities also showed the importance of the incidence of households with children. The most striking finding in this regard was the rank reversal of Hispanic and White families; Hispanics ranked slightly above Whites when household production was added to expenditures.

Including household production will lower the observed inequality in the distribution of a measure of consumption. Our assertion is based on analyzing the addition of household production to a standard measure of consumption expenditures. But, it is likely to hold for any other reasonable consumption metric. Our estimates show that, for consumer units in the bottom half of the standard expenditure distribution, the average value of household production is roughly as large as that of expenditures. For those in the top half, the average value of household production shows only a small variation across expenditure deciles compared with the sharp gradient of average expenditures. Thus, the relatively large size of household production and its somewhat equitable distribution across the expenditure distribution ensures that its addition will increase the expenditure shares[74] of those on the lower rungs of the distribution. This shift will be reflected in the new distribution, displaying a lower inequality than the prior distribution.

Two opposing forces affect the change in inequality. First, since the mean value of household production is notably higher for those with children than those without, adding home production to expenditures widens the intergroup inequality between the two groups in the augmented measure compared to the standard estimate of expenditures. However, because of the lesser inequality in the distribution of household production compared to the distribution of spending, the inequality in the augmented measure within both groups will be lower than in the standard measure. Irrespective of whether we use the BEA or our definition of household production, we found that the reduction in within-group inequality overwhelms the increase in between-group inequality.

The extent of the reduction in inequality is larger with the BEA definition than ours. The difference is primarily because our measure contributes more to inequality in augmented expenditures than the BEA measure. Assessing the contributions made by the components of household production to inequality revealed that the difference stems from childcare. As noted above, our childcare measure is broader than the BEA measure because we include supervisory care, and hence the share of childcare in augmented expenditure is higher under our definition. In turn, the higher share of childcare results in a bigger contribution to inequality.

Finally, we should note that the analysis of the matched data we provide illustrates the types of uses it has. Other assumptions and techniques can be applied, as in any scholarly research. For example, some of the findings reported here regarding subgroup disparities and Gini coefficients may be changed substantially if an equivalence scale were applied. We have also not estimated confidence intervals for the statistics reported in this chapter. Such an exercise may end up altering some of the inferences that we have made, e.g., regarding subgroup differentials.[75] Further, our results are predicated on our assumptions in converting the imputed values from a single weekday diary and single weekend-day diary to form weekly and monthly values. They also depend on our premises in aggregating from the imputed individual-level to household-level values. Alternative hypotheses and techniques could be deployed to

[74] For example, the share of each decile in the aggregate value of the new expenditures that includes the value of household production.

[75] A technical question here is the appropriateness of using the CE replicate weights with variables imputed via statistical matching. We cannot provide estimated imputation errors associated with our statistical matching. On the other hand, imputation errors can be accounted for by using the values imputed by the MI method. A proper investigation of this issue falls outside the scope of our report.

address these questions as well. We hope the matched data and our initial analysis spur further research on these topics.

6 Conclusion and Recommendations

The purpose of this chapter is to outline our recommendations for incorporating home production into a broader measure of consumption. We should note at the outset that, if such a measure is to be used for analyzing poverty, thresholds must be adjusted to account for home production needs in order to maintain consistency between the definitions of needs and resources.[76] Our recommendations do not address the issue of such adjustments as that would need to be based on separate research. Instead, we focus on recommendations based on the analysis and conclusions outlined in Chapters 2–4. Below, we present them in the same order as the chapters.

6.1 Scope of home production

Recommendation 1.1. The definition of home production should be expanded to include supervisory care of household children under 13. The measure of supervisory care in the ATUS (secondary care) approximates this form of care provisioning in the home.

Our estimates show that, on average, consumer units with children under 18 spend 87 and 175 hours per month on active and supervisory childcare, respectively. Because we assign to supervisory care only one-half of the rate we assign to active care, their monetary values are almost identical at just under $1,500 per month. Even so, the monetary value of supervisory care is about 27 percent of the total average market expenditures of households with children. Ignoring this would be neglecting a sizeable part of caregiving services used by the household.

Recommendation 1.2. Care received from people outside the consumer unit without payment (i.e., care given by nonhousehold relatives and nonrelatives) should be included in the definition of home production. In addition, data collection efforts should be enhanced for gathering information on nonmarket care received from nonhousehold members by older children (5 to 13 years) and disabled or sick individuals under 50 years living in households.

We estimated the care children under 6, and older adults (over 50) received from nonhousehold relatives and nonrelatives. Among consumer units with children under 6, we found that roughly 12 percent (2.1 million families) benefitted from the care given by nonhousehold people. The magnitude of the care received was quite substantial for recipient households. We estimated that the average monthly help for

[76] This issue of consistency has been discussed in detail in Citro and Michael (1995). The Levy Institute Measure of Time and Consumption Poverty that has been estimated for a variety of countries represents a potential approach to developing such a measure. More information is available at: https://www.levyinstitute.org/research/the-levy-institute-measure-of-time-and-income-poverty

these households was 90 hours with a monetary value of $1,500 or about 27 percent of the average monthly market expenditures of the recipient units. Among consumer units with a person over 51, about 4 percent of consumer units, or about 2.9 million households, received care from people outside the household. For the beneficiaries, the help received was sizeable in terms of time (about 106 hours per month) and as a proportion of their monthly market expenditures (40 percent). Given the omission of care received by people outside the age range of our estimates (individuals between 6 and 50 years) as well as the limited scope of care activities covered in the data sources for our calculations discussed in Chapter 4, our estimate is a lower-bound estimate of this type of caregiving.

6.2 Valuation of home production

Recommendation 2.1. At this point in time, input valuation based on replacement cost using specialist wages is the best available option for valuation of non-market work. Activities should be disaggregated where feasible and appropriate, with attention to joint production. Active childcare should be treated as a relatively skilled task that builds on child-specific knowledge and relationship continuity to yield important developmental consequences. Supervisory care of young children should be recognized as a significant constraint on parental time that imposes costs. Valuation of supervisory care for children under 13 should take into account engagement in simultaneous activities; domestic work undertaken alongside supervisory care should be valued more highly than that without. We recommend valuation of supervisory care at approximately one half the value of active care (typically less than a babysitter's wage). Further, geographical variation in replacement costs should be accounted for in the valuation by, e.g., using wage rates differentiated by metro areas.

6.3 Imputations from the ATUS

Recommendation 3.1. Considering that the ATUS is representative for quarters, we recommend that each quarterly CE Interview or Diary sample should be treated as an independent recipient file in the imputation rather than, for example, pooling all four quarters of the Diary sample for a given year as a single recipient file. This will ensure, at the very least, that imputed values for consumer units in each quarter are representative of the full-year data.

Recommendation 3.2. The weekday and weekend 24-hour diaries in the ATUS for a given year should be considered separate donor files, given the significant differences in time-use patterns between weekends and weekdays. Therefore, individuals over 14 years in the quarterly CE or Diary sample should have separate imputations from the weekday and weekend diaries, which would provide a better representation of weekly time-use patterns. Hours spent on home production per week can be calculated as a weighted average of the imputed weekday and weekend values.

Recommendation 3.3. Home production times should be disaggregated for each record in the synthetic (imputed) CE files. The categories to be created include cooking, other housework (including shopping), adult care, active childcare, supervisory childcare overlapping with other home production, and supervisory childcare that does not overlap with other home production. This may provide enough disaggregation for future research, without compromising on the quality of the data imputation. To avoid double counting in calculating the total time spent on household production, the time during which supervisory care coincides with other household production should be included only in *one* of the following categories: supervisory care overlapping with other household production, cooking, other housework, or adult care.

Recommendation 3.4. Supervisory childcare time for individuals in multi-adult households should be adjusted to avoid potential double-counting. Adjustments should be made on parameters differentiated by gender and the number of likely caregivers estimated from the ATUS. Better information is required to understand the dynamics of supervisory care provision in households with multiple caregivers and care receivers. Unfortunately, this cannot be obtained under the current ATUS sampling strategy of collecting time diaries from only one respondent per household.

Recommendation 3.5. For each record, the home production categories mentioned in Recommendation 3.3 above should be imputed according to the three imputation methods described in Chapter 3. The methods are regression prediction (RP), multiple imputations (MI) based on regression models that account for censoring, and constrained statistical matching (SM). The availability of imputations based on alternative methods will offer the users some flexibility in exercising their methodological preferences.

Recommendation 3.6. Consumer unit-level values of home-production times according to the categories identified in Recommendation 3.3 should be constructed by adding up the times for individuals within the consumer unit. This operation should be conducted for the three imputation methods mentioned in Recommendation 3.5.

Recommendation 3.7. The monetary values of household production times described in the previous recommendation should be constructed according to our preferred method and the generalist wage method outlined in Recommendation 2.1 above. Researchers would thus have the option of testing the sensitivity of their findings to the valuation principle. Providing individual-level time-use data may be advisable, giving researchers further flexibility to test hypotheses and sensitivity to the valuation assumptions.

6.4 Imputations of care received from outside the consumer unit

Recommendation 4.1. Nonmarket help received by members of the consumer unit from people outside their household (relatives and nonrelatives) should be imputed from the Health and Retirement Survey (HRS) and Early Childhood Program Participation (ECPP) survey. In our view, the HRS and ECPP are the best, though far from ideal, sources currently available for the incidence and duration of these forms of care. The hours of adult and childcare received at the consumer-unit level can be derived by adding the hours of care received by their members. Further, the hours should be converted into monetary values using the hourly rate applied to adult care and active childcare hours provided by household members.

6.5 Recommendations for the use of synthetic CE public-use files

In theory, imputations are only as good as the strength of the conditional independence assumption (CIA). A direct test of CIA is generally impossible because the key variables we seek to study jointly are not available in the same sample, hence the need for imputation. Nevertheless, in our case, information on consumption expenditures and time spent on household production is collected in the Panel Study of Income Dynamics (PSID). While the expenditure definitions and time-use categories in the PSID do not match those in the CE and ATUS, we believe that PSID is a reasonably good (and only, so far) data platform to conduct a test of the CIA. As we discussed in Appendix A of Chapter 3, the results from our testing strongly support the CIA.[77] Indirectly, the evidence offered solid methodological support for the joint analysis of imputed time-use variables and observed consumption expenditures in the CE files. It is also worth noting that reputed academic journals publish papers that include imputed data as though it were actual data in statistical modeling without much concern for the CIA or adjustment for imputation errors (see Chapter 3 for some examples).

We have provided detailed information regarding the quality of the imputations in the report and online-data appendix. Overall, statistical matching (SM) performed as well and, in some cases, somewhat better than the prediction methods, i.e., conditional mean (RP) and multiple imputations (MI), in replicating the average values of time spent on household production by men and women in various population subgroups from the American Time Use Survey (ATUS). Regarding medians of the same subgroups, SM produced imputed values that are, in general, closer to their ATUS counterparts than the other two

[77] As noted in Appendix A, Chapter 3, our validation exercise was conducted using total consumption expenditures. However, the method outlined there can be deployed to study the relationship between a specific broad category of expenditure, say food expenditures, and time-use variables using the PSID, bearing in mind the limitations of the data with respect to the categories. Advanced users seeking to explore such relationships using the imputed household production variables in the CE samples may want to conduct supplementary analysis using the PSID as a robustness check.

techniques. Further, we found that the within-subgroup distribution of hours in the original data was also transferred to the synthetic CE files in a much more reliable manner by SM than by the other two imputation strategies. Thus, we recommend using the variables imputed by SM for descriptive analysis, such as calculating means, medians, and inequality measures (e.g., the Gini coefficient).

A caveat applies to our quality assessment. Suppose the research focus is at the consumer-unit level rather than the individual level. In that case, the researcher should bear in mind that our evaluation was conducted at the individual level (e.g., looking at the closeness of the imputed and observed median value of hours spent on cooking by employed men with children). We cannot perform quality assessments at the household level because the ATUS does not collect time diaries from everyone above 14 years in the household. However, since the household-level aggregate would result from the summation of individuals' values in the home, we should expect good-quality imputations at the individual level to translate into good-quality imputations at the household level.

Further, imputations generally tend to be of lower quality for smaller population subgroups. This caveat holds especially if the group, on average, does not engage much in household production (e.g., men between 15 and 24 years). Users should, therefore, exercise caution in splitting the imputed data into finer subgroups. We would suggest a preliminary inspection of the ATUS data to ascertain the incidence of household production (or subcategories of household production such as cooking) in the subgroups of their interest to avoid misleading results.

The main drawback of the SM and RP strategies is their inability to account for imputation errors in conducting statistical inference in the synthetic file, e.g., testing the equality of means for hours spent on childcare between two subgroups in the CE Interview sample from the second quarter of 2019. Specialized procedures are available in statistical software such as SAS and Stata to analyze multiple-imputed data (e.g., PROC MIANALYZE in SAS or `mi estimate` in Stata). Users of the multiple-imputed home-production values in the CE synthetic files should consult the documentation of their statistical software in choosing the methods appropriate for such data.

Since SM is better at replicating the original ATUS data than MI, there is a paradox. While the MI method allows theoretically justifiable statistical inference, the statistically matched values may yield better-quality parameter estimates. Therefore, we recommend the analysis of interest using both sets of imputed values and comparing the results before stating solid conclusions. For example, consider using imputed hours of childcare of employed mothers as the dependent variable and some measure of market consumption as the independent variable, along with control variables (e.g., marital status, age, number of children, etc.) in a regression model. First, we would estimate the model using the imputed values of SM and MI separately. Then, suppose a comparison shows that the size and statistical significance of the coefficient on market consumption is the same irrespective of the set of imputed values used in the estimation. In that case, we can be reasonably confident about the estimated relationship. However, caution is required if there are incongruences in the parameter estimate or significance. And the researcher should exercise their judgment regarding the substantive importance of the incongruence.

Appendix A Validation exercise using the PSID

Background

The fundamental assumption behind imputation methods is the Conditional Independence Assumption (CIA). Basically, the premise is that, once we control for a set of relevant characteristics, the distribution of variables that cannot be observed jointly and require imputation is as good as independent. The assumption justifies the creation of a synthetic joint dataset by taking random draws from the conditional empirical (with or without replacement) or theoretical distribution.[78]

In the current project that seeks to impute time spent on household production by individuals in the CE using the ATUS, the CIA implies that time spent on household production activities by individuals is conditionally independent of household consumption expenditures. While this assumption is, in principle, untestable, we use the Panel Study of Income Dynamics (PSID), which has data on both time-use and consumption expenditures, to explore the level of correlation between time use and consumption as a validation exercise. Of course, the scope and nature of PSID are different from those of the ATUS and CE. However, several studies have found that the PSID provides good approximations of aggregate expenditure and time use, comparable to the CE and the ATUS.[79] Thus, findings using this dataset may prove helpful as evidence in favor of or against the CIA.

Data

We use data from the PSID for the year 2019. This dataset has rich information on household and family characteristics, expenditure patterns, and detailed information on usual time-use activities and work history for the head of the household and his (her) spouse.

It should be emphasized that the definition of time use in the PSID is different from that captured in the ATUS. In the PSID, respondents are the reference people and spouses in married-couple households and the reference person if that person is single. The ATUS collects time-use information from one member, 15 years or older, of each household in the sample. Also, the PSID collects information via

[78] Rubin (1986) proposed a method that does not assume CIA. Instead, he proposes to combine the donor and recipient files using assumptions about the value of the unobserved partial correlation between the variables of interest (e.g., time spent on household production and consumption expenditures). The initial value of the partial correlation is estimated from the parameters of separate linear regressions using common regressors in the donor and recipient files (e.g. OLS of consumption expenditures on a set of characteristics in the CE and OLS of time spent on household production on the same set of characteristics). In contrast, CIA assumes the value to be equal to zero. However, both approaches involve placing restriction on the partial correlation coefficient. Our exercise in this appendix uses a particular method and data where both variables are observed to assess the appropriateness of our assumption of CIA.

[79] See Insolera et al. (2019) for a comparison of PSID and ATUS. Andreski et al. (2014) provide a comparison of consumption expenditures in PSID and CE.

recall on the amount of time per week devoted to eight non-work activities: housework, personal care, shopping, childcare, adult care, education, volunteering, and leisure. In contrast, the ATUS collects information using a time diary that records the previous day's activities. Using the recall approach as in the PSID does not consider activities overlapping, and adding the time spent on all activities may suggest individuals engage in more than 168 hours of activities per week. Despite this limitation, which should not have serious implications for our purposes here, we define the time spent on household production as the unadjusted sum of the time spent on housework, shopping, childcare, and adult care. We converted weekly hours into annual hours via multiplication by 52.

We estimate annual household consumption from expenditure aggregates constructed and readily available in the PSID 2019 datafile. Our approach is similar to that followed by Attanasio and Pistaferri (2014) and Fisher et al. (2020) to the treatment of mortgage and auto expenditures:

> Total Household Consumption = Food expenditure + Housing expenditure + Transportation expenditure + Education expenditure + Childcare expenditure + Health care expenditure + Computing expenditure + Clothing expenditure + Travel expenditure + Other recreation expenditure – Mortgage payments + Rent (if renting) + Imputed rent @ 6 % of home value (if owning).

We use the log of annual consumption in our analysis below.

Methodology

Consider three sets of variables of interest, consumption (X), time spent on household production (Y), and variables that affect both or, in the imputation context, common variables available in both surveys that can be used in the imputation (Z). Then, if the CIA holds, the distribution of X|Z and Y|Z are independent of each other, which implies:

$$f_{XY|Z} = f_{X|Z} * f_{Y|Z}$$

To test this assumption, we propose three complementary approaches. The first approach focuses on estimating the correlation between consumption and household production. If the two variables are independent, we would expect they are also linearly independent, and the correlation between them is zero. Thus, a simple test for the assumption is needed to estimate the correlation between household production and expenditure components.

$$\rho = Corr(X, Y)$$

To avoid problems related to the skewness of consumption data and the large number of zeroes in household production data, instead of estimating the correlation between the levels of the variables, we estimate the correlation between the cumulative density functions (CDF) of these variables:

$$\rho = Corr(F_X, F_y)$$

Using this approach would be equivalent to using ranked correlations, similar to the Spearman rank correlation, which is robust to any transformation of the underlying data.

To control for other factors Z, we define a set of clusters C_Z^J, which identifies J groups of individuals with similar characteristics, Z. These clusters are determined using the partition-mean-clusters algorithm. Thus, for a set of clusters C_Z^J, the average correlation will be defined as:

$$\rho_{CJ} = \sum_{j=1}^{J} sh_j \, Corr\left(F_{X|C^J=j}, F_{y|C^J=j}\right),$$

where sh_j is the proportion of observations in cluster $C^J = j$, and $F_{y|C^J=j}$ is the CDF within that cluster. If the correlation ρ_{CJ} is close to zero when using a particular set of clusters C_Z^J, we would conclude that the CIA assumption holds. Because our statistical matching approach uses clusters as the first step in identifying statistical twins, the evidence from this exercise could be considered to be appropriate for testing CIA in the matching context.

Our second approach is geared toward the technique of imputation by prediction and focuses on the correlation between residuals. The idea springs from the attempt to decompose the variation of the variables of interest—household production and consumption—into an explained component (which depends on Z) and an unexplained component. We can model these two components as some linear or nonlinear function of Z as follows:

$$X = f(\theta, Z) + e_x$$

$$Y = g(\theta, Z) + e_y$$

where, for simplicity, we assume that the error components e_x and e_y have a linear impact on the outcome, and $f(\theta, Z)$ and $g(\theta, Z)$ indicates we are using a sufficiently flexible functional form to avoid misspecification problems.

In this framework, e_x and e_y represent the distribution of X and Y after conditioning or absorbing the impact of all other characteristics Z. Thus, a direct test for the CIA assumption in this framework is to analyze the correlation between these residual errors:

$$\rho_r = corr\left(e_x, e_y\right)$$

For our present analysis, we will use a log-linear model to estimate the residuals for total-household expenditure. Given a large number of zeroes for household production data, we will use a Tobit model and obtain generalized residuals. Because of the nature of the procedure, the results we obtain here will be appropriate as evidence for the CIA in the imputation-by-prediction approach.

Finally, we also consider a modification of the second approach just outlined. The main difference is that, rather than using a two-step approach to verify the CIA, we directly add the variable of interest to the model specification to confirm if any statistically significant relationship exists between consumption and household production.

Specifically, we propose to estimate the following models:

$$X = f(\theta, Z) + \beta Y + e_x$$

$$Y = g(\theta, Z) + \gamma X + e_y$$

Here, we can test directly if the CIA holds after controlling for Z by examining the coefficients β and γ.

We now describe the clusters and control variables deployed in the testing of CIA in the prediction context. We obtain the optimal clusters for testing CIA using conditional correlations via the following procedure:

i. Using all observations in a strict strata group (say, women with children), apply principal component analysis to the complete set of characteristics Z and obtain all principal components with an eigenvalue larger than 1.
ii. Use the principal component to identify J clusters C^J via the mean-partition algorithm.
iii. Given the random nature of the cluster analysis, this step is repeated multiple times, and the cluster that maximizes the Calinski-Harabasz statistic is chosen as the optimal cluster.
iv. Steps ii and iii are used to create clusters that identify 10, 20, …, 100 groups.
v. Steps i–iv are repeated for each subgroup segmented by gender and the presence of children (under 18) in the household.
vi. Finally, all clusters are combined with gender and the presence of children to identify groups for the analysis. Observations are classified into up to 400 groups.

For testing CIA using residual correlation and direct-regression methods, we select a set of variables common to the ATUS and CE, also available in the PSID survey. These variables are:[80]

> Individual-level: gender, age, years of education, a dummy for being employed, a dummy for self-employment, average hours of work per week last year, and a dummy for spouse present;
>
> Spouse level (if relevant): age, years of education, a dummy for being employed, a dummy for self-employment, and average weekly hours of employment last year.

[80] The control variables used here overlap with those used in Attanasio and Pistaferri (2014) for their imputation of consumption expenditures in the PSID. In addition, we tested a model with variables related to the earnings of the head and spouse (if applicable) as in Crossley et al. (2020). The results were not substantially different.

Household:

Dummy for home ownership and number of rooms in the housing unit.

Family Characteristics:

Average years of education of all adult members, age of the youngest child, number of children by age group (0–1; 2–5; 6–12; 13–17), number of adults by gender in various age groups (18–24; 25–44; 45–64; 65+), number of employed adults, and a dummy for a child (0–17 years) in the household.

Except for gender and the presence of a child in the home, all other variables are used as is in the set of control variables Z. Given the sizeable difference in the time spent on household production between men and women and between individuals with and without children in their household, a combination of these two variables is used as a strict strata variable for the average conditional correlation approach,[81] or used interacted with the rest of the variables, for the regression approach.

Results

We start with the analysis of the CIA using the conditional correlation approach. In all cases, the main focus should be the rank correlation between total consumption and time spent on household production. As supplementary information, we also present the changes in the conditional correlation for the components of household production that can be identified in the PSID.

Figure 1 relates the changes in the conditional correlation between household production and total household expenditure. The first point to consider is that, when no controls are used, the rank correlation between spending and time spent on household production is just above 0.12. Once we control for at least gender and the presence of a child, it drops drastically to 0.038, a level considered small and with few economic implications. Interestingly, using the sub-clusters does not further reduce the average conditional correlation, which remains low, around 0.05.

[81] Using these as strict strata variables means that we will never form a cluster that will include both men and women or include men in households with and without children or women in households with and without children.

Appendix Figure A-1 Conditional correlation: total household production

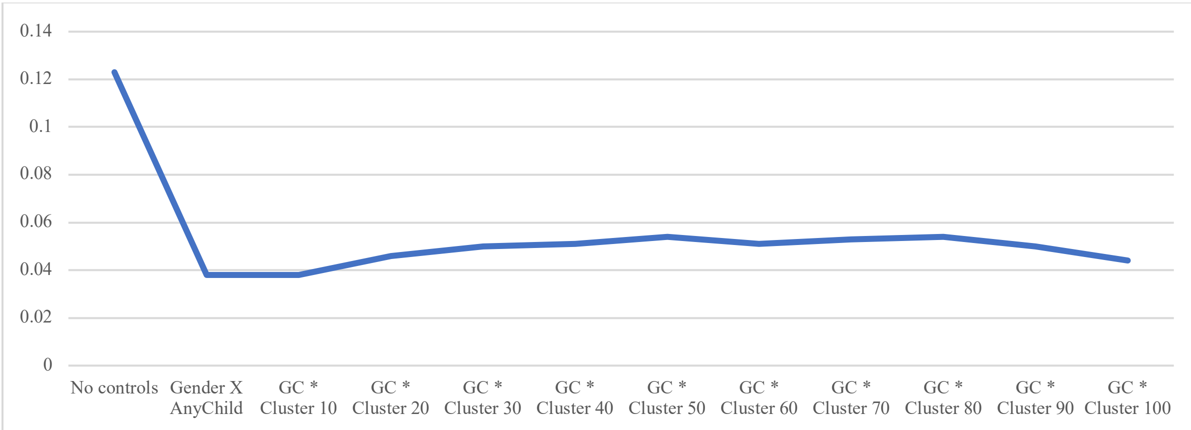
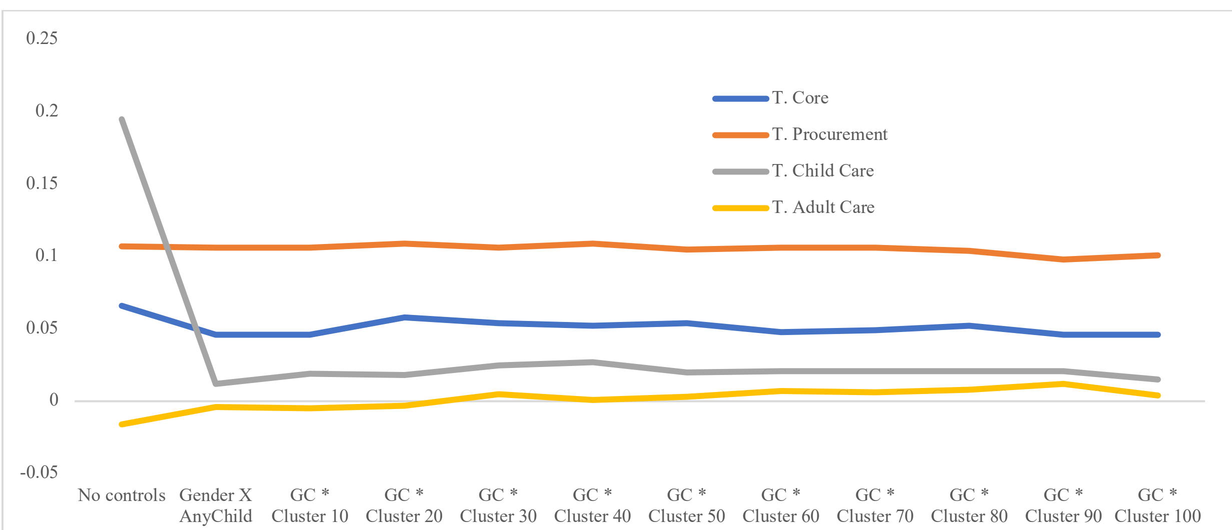


Figure A2 presents a similar exercise as Figure A1, but where correlations are estimated using the components of household production: housework ("core"), shopping ("procurement"), childcare, and adult care. Since childcare is often the focus of the literature on the care economy, it is interesting to note that, when no controls are included, there is a strong correlation of almost 0.2. Like total household production, accounting for gender and the presence of children alone absorbs almost completely the correlation between childcare and consumption to negligible levels. In contrast, time spent on adult care seems uncorrelated with household consumption, even when we do not control for any characteristics.

Appendix Figure A-2 Conditional correlation: components of household production



The correlation of consumption with time spent on core activities is low, less than 0.1. Moreover, it reduces considerably after controlling for gender and the presence of children and accounting for at least ten groups of subclusters. The correlation, however, does not fall further, regardless of the number of clusters used. The last component of interest is procurement (mainly shopping). In contrast with core activities and adult care, this variable shows a notable correlation with expenditure (0.1) regardless of the number of clusters used. The stable behavior, however, may make economic sense: households with higher expenditures may also spend more time shopping, which explains the observed correlation. However, while this correlation may be considered significant, we should place the time dimension in context. On average, people spend only 4 hours per week on procurement, and less than 6 percent of people spend over 10 hours on the same.

Overall, PSID evidence suggests that the CIA assumption holds for total household production, core activities, adult care, and childcare. While the evidence is perhaps less strong for procurement, one should consider that the average conditional correlations presented here only represent an upper bound of the underlying conditional correlation across variables, as we do not fully control for differences in characteristics within a particular cluster. In the statistical matching approach, we address this by using the rank implied by the propensity score to further aid in identifying statistical twins.

We now turn to the results of testing CIA using regression models. As already described, the residual correlation focuses on obtaining the residuals (or generalized residuals) from estimating a model with household production or household consumption as the dependent variable, controlling for all common variables Z. We use an OLS that specifies the log of household consumption as a function of control variables mentioned in the previous section, all interacted with gender and the presence of child dummies. We use a Tobit model for household production and obtain the generalized residuals from the model.

The underlying models are not reported here because of the large number of explanatory variables but are available upon request. The household expenditure model has an R-squared of 0.484 and uses 105 explanatory variables. The time-use Tobit models have a lower, pseudo R-squared. The highest is observed for time spent on childcare (0.118), and the lowest is for procurement (0.018). The pseudo R-squared for total household production is 0.048.

Once the residuals from the model are obtained, we can construct the corresponding correlations to test for the CIA assumption. Table A1 provides these correlations between total household expenditure and time spent on household production. Similar to before, we report the correlations from three models. One without controls, one that controls for gender and the presence of children, and one with full interaction (GC X Z).

Appendix Table A-1 Residual correlation

	Household Production	Core	Procurement	Childcare	Adult Care
No controls	0.0515	0.0367	0.0707	0.1443	-0.0168
Gender and Child Presence	-0.0419	0.0316	0.0909	-0.0240	-0.0054
GC X Z	0.0260	0.0270	0.0560	0.0160	0.0150

The results align with the previous approach's findings, although there are some differences. At the outset, the correlation coefficients without any controls are relatively low for all household production categories, except perhaps for childcare (0.14). Controlling for gender and child presence seems to reverse the sign of the correlation coefficient for household production and childcare. In fact, the correlation coefficient drops considerably for childcare to –0.02. Controlling for other factors has almost no effect on core activities or procurement. However, for total household production, we observe a sharp decline. As in the previous case, procurement still correlates highest with expenditure compared to other time components.

Finally, we present the results from the direct regression approach of testing CIA—which includes the variable of interest in the model specification—and verify that the effects are small and non-significant. In Table A2, we report only the coefficient for the variable of interest and its standard error. Similar to the two previous cases, we report the estimated coefficient for a model without controls, except for the log of household expenditure, a model that controls for gender and the presence of a child, and the fully interacted model. Once again, the results are pretty consistent with the previously presented evidence.

Appendix Table A-2 Conditional correlation coefficient of log(hh Exp)

Coefficient of log (HHexp)	Household Production	Core	Procurement	Childcare	Adult Care
No controls	3.712***	0.742**	0.599***	19.24***	-2.551
	(0.710)	(0.233)	(0.107)	(1.360)	(1.675)
Gender and Child Presence	-2.707***	0.622**	0.775***	-2.735*	-0.830
	(0.674)	(0.229)	(0.108)	(1.215)	(1.706)
GC X Z	0.883	0.564	0.815***	0.967	0.903
	(0.820)	(0.293)	(0.133)	(1.513)	(2.188)

Note: Standard errors in parenthesis. *** - significant at 1%, ** - significant at 5%.

In the models with no controls, the impact of household consumption is statistically significant (except for time spent on adult care), especially in the models for the time spent on childcare and total household production. Controlling for gender and child presence substantially reduces the coefficient on childcare and reverses its sign. Nevertheless, the effect remains significant. We also observe that the impact on household production remains significant but with a negative sign. However, once we use the fully interacted model, the statistical significance of the relationship between household expenditure, total time on household production, and time spent on childcare disappears. While the effects for procurement and core activities remain statistically significant, the magnitudes of the effects suggest that they may not be economically significant. Specifically, the model predicts an increase of 0.07 and 0.08 hours (less than five minutes) in time spent on core activities and procurement for a 10 percent change in expenditure.

To summarize, the evidence suggests that the CIA assumption holds, at least once demographic differences and family structure characteristics are controlled. Moreover, these results are consistent across all the methodologies for testing for the CIA we presented here. But, our testing is with total expenditures. The outcomes may not be so encouraging if the relationship is examined for subcategories of expenditure and time use (e.g., food expenditures and time spent on household production).

Appendix B Comparison of the results from two algorithms of multiple imputation

As described in the main text, we use two different algorithms for the multiple imputation (MI) approach. First, when using the statistical software Stata, we use an interval regression model, a generalization of the Tobit-censored regression model, to identify the data-generating process of time-use activities. The model has sound econometric properties and is meant to correctly address the presence of excess zeroes in data, as is the case for time use.

Due to software limitations, when using the statistical software SAS, we use a fully conditional specification method (FCS) based on a linear regression (LR) model. Predictions of the model are censored out after all imputations are created. Because this approach uses an LR model for a limited dependent variable with a large number of zeroes, we expect the coefficients and predictions to be biased after the censored correction. SAS does not have a Tobit model for multiple imputation procedures.

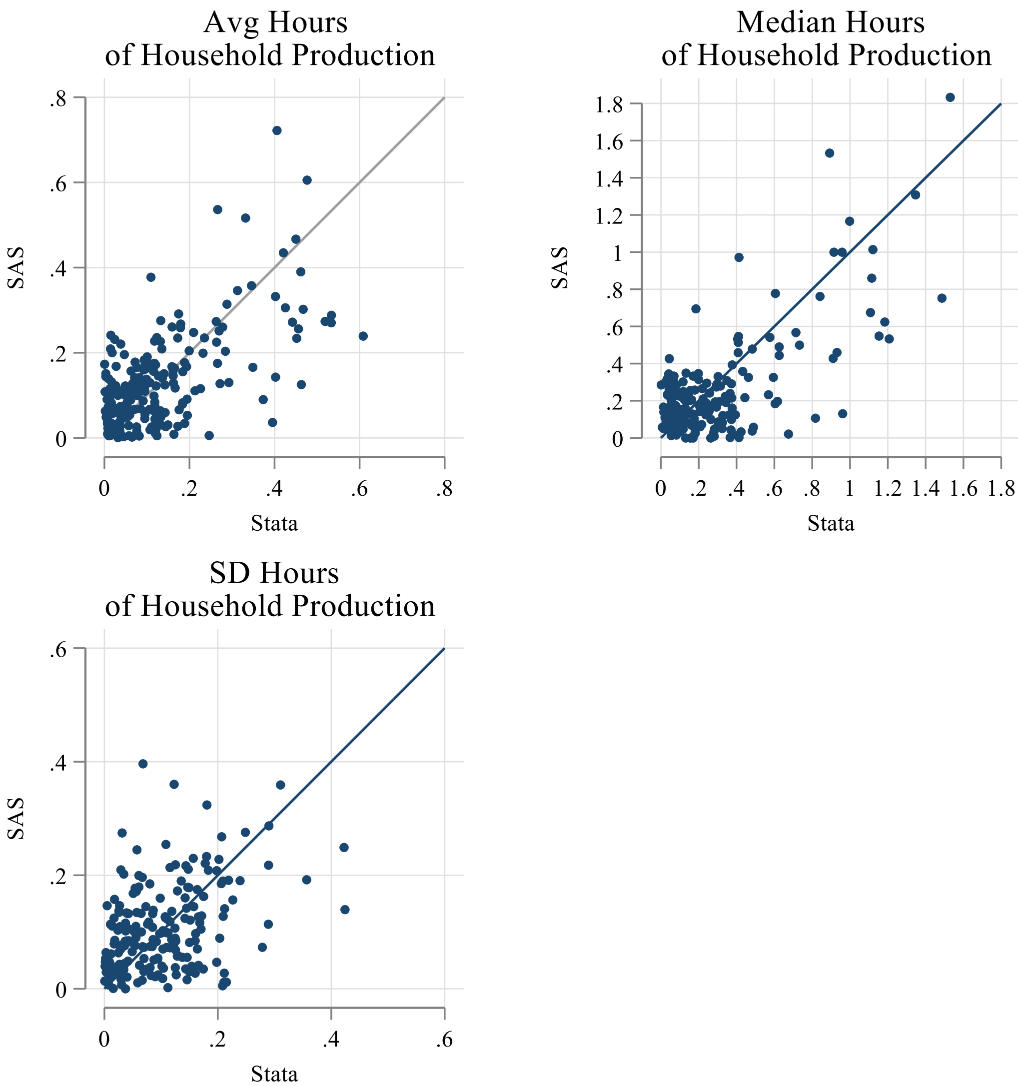
To compare the performance of the procedures, we estimate the absolute relative difference of mean, median, and standard deviation for imputed total hours of household production for various subgroups and compare them across the Stata (Tobit) and SAS(LR-FCS) procedures. We concentrate on household production over weekends, focusing on a single replicate. The results for weekdays are similar.

Figure B1 provides a scatter plot of the absolute relative differences for mean, median, and standard deviation. If the SAS procedure were superior, we would see that the scatterplot of points to lean below and to the right of the main diagonal. In contrast, if the Stata procedure is superior, we should see the opposite, with the scatterplot of points leaning upward and to the right of the main diagonal.

What we observe, however, is that, despite the theoretical advantages of the Tobit model that Stata has on characterizing the conditional distribution of time, given the large presence of zeroes, there is no clear advantage for the Stata results. The cloud of points distribution suggests that both approaches have advantages and disadvantages across different subgroups.[82]

[82] The cloud of points represents absolute percentage difference between the ATUS subgroup statistic and the imputed data. The closer the point is to zero, the better was the imputation to carry over the distribution of time use. Points below the 45-degree line suggest SAS FCS procedure produces better imputed values, whereas points above the line suggest Stata-censored regression imputation to perform better.

Appendix Figure B-1 Comparison of SAS and Stata methods of multiple imputation



References

Aassve, A., Meroni, E., & Pronzato, C. (2012). "Grandparenting and Childbearing in the Extended Family." *European Journal of Population*, *28*(4), 499–518.

Andreski, P., Li, G., Samancioglu, M. Z., & Schoeni, R. (2014). "Estimates of Annual Consumption Expenditures and Its Major Components in the PSID in Comparison to the CE," *American Economic Review*, *104*(5): 132–135.

Antonopoulos, R. (2008). *The Unpaid Care Work-Paid Work Connection.*Annandale-on-Hudson, NY: Levy Economics Institute of Bard College. Working Paper No. 541.

Armstrong, G., Cho, C., Garner, T. I., Matsumoto, B., Munoz, J., & Schild, J. (2022). "Building a Consumption Poverty Measure: Initial Results Following Recommendations of a Federal Interagency Working Group. *AEA Papers and Proceedings*, *112*: 335–339. https://doi.org/10.1257/pandp.20221041

Attanasio, O., & Pistaferri, L. (2014). "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," *American Economic Review*, *104*(5): 122–126.

Brady, M. (2016). "Gluing, Catching and Connecting: How Informal Childcare Strengthens Single Mothers' Employment Trajectories," *Work, Employment & Society*, *30*(5): 821–37.

Bridgman, B., Craig, A., & Kanal, D. (2022). "Accounting for Household Production in the National Accounts: An Update 1965–2020," *Survey of Current Business*. https://apps.bea.gov/scb/2022/02-february/0222-household-production.htm

Cardia, E., & Ng, S. (2003). Intergenerational Time Transfers and Childcare," *Review of Economic Dynamics*, *6*(2): 431–54. https://doi.org/10.1016/S1094-2025(03)00009-7

Casper, L., M. (1996). *Who's Minding Our Preschoolers?* (No. P70-53; Household Economic Studies). U.S. Census Bureau. https://www.census.gov/library/publications/1996/demo/p70-53.html

Citro, C. F. and Michael, Robert T. 1995. *Measuring Poverty: A New Approach*. National Academy Press.

Coleman, M. (1998). "Homemaker as Worker in the United States." *Challenge* 41(6): 75–87.

Collins, A. & Carlson, B. (1998). *Childcare by kith and kin: Supporting family, friends and neighbors caring for children* (No. 5; Issue Brief). National Center for Children in Poverty. https://www.researchconnections.org/childcare/resources/2157

Crossley, T., Levell, P., & Poupakis, S. (2020). "Regression with an Imputed Dependent Variable" (No. 20/25; Working Paper). London: Institute for Fiscal Studies.

Cui, J. & Natzke, L. (2021). *Early Childhood Program Participation: 2019*. (NCES 2020-075REV), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved January 28, 2023 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020075REV.

Enders, C. K. (2022). *Applied missing data analysis* (Second Edition). The Guilford Press.

Ettner, S. L. (1995). "The Impact of "Parent Care" on Female Labor Supply Decisions," *Demography*, *32*(1): 63–80. https://doi.org/10.2307/2061897

Fisher, G. G. & Ryan, L. H. (2018). Overview of the Health and Retirement Study and Introduction to the Special Issue. *Work, Aging and Retirement*, *4*(1), 1–9.

Fisher, J. D., Johnson, D. S., Smeeding, T. M., & Thompson, J. P. (2020). "Estimating the Marginal Propensity to Consume Using the Distributions of Income, Consumption, and Wealth," *Journal of Macroeconomics*, *65*, 103218.

Folbre, N., Murray-Close, M., & Suh, J. (2018). "Equivalence Scales for Extended Income in the U.S. *Review of Economics of the Household*, *16*(2): 189–227. https://doi.org/10.1007/s11150-017-9387-8

Folbre, Nancy. (2009). "Time Use and Living Standards." *Social Indicators Research* 93(1): 77–83.

Folbre, N., Yoon, J., Finnoff, K., & Fuligni, A. S. (2005). "By What Measure? Family Time Devoted to Children in the United States." *Demography* 42 (2): 373–90.

Frazis, H. & Stewart, J. (2011). How Does Household Production Affect Measured Income Inequality? Journal of Population Economics, 24(1): 3–22.

Fuchs, V. R. (1986). "His and Hers: Gender differences in Work and Income, 1959-1979." *Journal of Labor Economics* 4(3, Part 2, S245-S272.

Garner, T. I., Matsumoto, B., Schild, J., Curtin, S. & Safir, A. (2023). "Developing a consumption measure, with examples of use for poverty and inequality analysis: a new research product from BLS," Monthly Labor Review, U.S. Bureau of Labor Statistics, April. https://doi.org/10.21916/mlr.2023.8

Gautham, L. & Folbre, N. (forthcoming). "Parental Expenditures of Time and Money on Children in the U.S.," *Review of Income and Wealth*.

Goldschmidt-Clermont, L. (1993). "Monetary Valuation of Non-Market Productive Time: Methodological Considerations." *Review of Income and Wealth* 39(4): 419–33.

Grigoryeva, A. (2017). Own Gender, Sibling's Gender, Parent's Gender: The Division of Elderly Parent Care among Adult Children. *American Sociological Review*, *82*(1): 116–46.

Hawkins, J., Habib, K. N. (2022). "A Multi-source Data Fusion Gramework for Joint Population, Expenditure, and Time use Synthesis". Transportation. https://doi.org/10.1007/s11116-022-10279-8

He, D. & McHenry, P. (2016). "Does Formal Employment Reduce Informal Caregiving?" *Health Economics*, *25*(7): 829–43. https://doi.org/10.1002/hec.3185

Herbst, C. M. (2023). "Childcare in the United States: Markets, Policy, and Evidence," *Journal of Policy Analysis and Management*, *42*(1): 255–304. https://doi.org/10.1002/pam.22436

Insolera, N., E., Johnson, D., S., & Simmert, B., A. (2019). *Evaluation of the Time Use data in PSID with comparisons to ATUS* (#19-02; Technical Series Paper). Institute for Social Research, University of Michigan.

Jenkins, S. P. & O'Leary, N. C. (1996). Household Income Plus Household Production: The Distribution of Extended Income in the U.K. Review of Income and Wealth, 42(4): 401–19.

Jokubauskaitė, S. & Schneebaum, A. (2021). "Assessing the Value of Household Work Based on Wages Demanded on Online Platforms for Substitutes," *Review of Economics of the Household* 20: 153–60.

Kneeland, H. (1929). Woman's Economic Contribution in the Home. *Annals of the American Academy of Political and Social Science*, *143*(May): 33–40.

Kum, H., & Masterson, T. N. (2010). Statistical matching using propensity scores: Theory and application to the analysis of the distribution of income and wealth. *Journal of Economic and Social Measurement*, *35*(3), 177–196.

Kuznets, S, Epstein, L. & Jenks, E. (1941). "National Income and Its Composition, 1919-1938, Volume II." New York: National Bureau of Economic Research (NBER). http://www.nber.org/books/kuzn41-3 (January 4, 2013).

Landefeld, J. S., Fraumeni, B. M., & Vojtech, C. M. (2009). "Accounting for Household Production: A Prototype Satellite Account Using the American Time Use Survey," *Review of Income and Wealth*, *55*(2): 205–25.

Landefeld, J. S., & McCulla, S. H. (2000). "Accounting for Nonmarket Household Production within a National Accounts Framework." *Review of Income and Wealth*, *46*(3): 289–307.

Laughlin, L. (2013). *Who's Minding the Kids? Childcare Arrangements: Spring 2011* (No. P70-135; Household Economic Studies). U.S. Census Bureau. https://www.census.gov/library/publications/2013/demo/p70-135.html

Maestas, N., Messel, M., & Truskinovsky, Y. (2020). "Caregiving and Labor Force Participation: New Evidence from the Survey of Income and Program Participation," *No. WI20-12). Madison, WI: Center for Financial Security University of Wisconsin Madison Retirement and Disability Research Center*.

Milkie, M. (2022). "Beyond Mothers' Time in Childcare: Worlds of Care and Connection in the Early Life Course." Manuscript, Department of Sociology, University of Toronto, under review at *Time and Society.*

Moussa, M. M. (2019). "The Relationship between Elder Care-giving and Labour Force Participation in the Context of Policies Addressing Population Ageing: A Review of Empirical Studies Published between 2006 and 2016," *Ageing & Society*, *39*(6): 1281–310. https://doi.org/10.1017/S0144686X18000053

Mullan, K. (2010). "Valuing Parental Childcare in the United Kingdom." *Feminist Economics* 16(3): 113–39.

National Research Council. (2005). *Beyond the Market: Designing Nonmarket Accounts for the United States* (K. G. Abraham & C. Mackie, Eds.). The National Academies Press. https://doi.org/10.17226/11181

Newman, D. A. (2014). "Missing Data: Five Practical Guidelines." *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Osuna, V. (2021). "Subsidising Formal Childcare Versus Grandmothers' Time: Which Policy is More Effective?" *Economics: Journal Articles*, *15*(1): 85–111. https://doi.org/10.1515/econ-2021-0007

Reid, M. G. (1934). *Economics of Household Production*. New York: J. Wiley & Sons.

Royston, P. (2007). Multiple Imputation of Missing Values: Further Update of Ice, with an Emphasis on Interval Censoring. *The Stata Journal: Promoting Communications on Statistics and Stata*, *7*(4), 445–464. https://doi.org/10.1177/1536867X0800700401

Rubin, D. B. (1986). "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations" *Journal of Business & Economic Statistics*, 4(1) (Jan., 1986): 87–94.

Shapiro, L. (2009). *Perfection Salad: Women and Cooking at the Turn of the Century*. University of California Press.

Suh, J. & Folbre, N. (2023). "The Responsibilities of Parental Childcare: Evidence from the American Time Use Survey**,"** paper presented at the meetings of the Allied Social Science Meetings, New Orleans, LA, January.

Suh, J. & N. Folbre. (2015). “Valuing Unpaid Childcare in the US: A Prototype Satellite Account Using the American Time Use Survey.” *Review of Income and Wealth*, April, 1–17.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*(3), 219–242.

Wheelock, J. & Jones, K. (2002). ‘Grandparents Are the Next Best Thing’: Informal Childcare for Working Parents in Urban Britain. *Journal of Social Policy*, *31*(3), 441–463. https://doi.org/10.1017/S0047279402006657

Wolff, E. N. & Zacharias, A. (2007). “The Levy Institute Measure of Economic Well-Being: United States, 1989–2001.” *Eastern Economic Journal,* 33(4): 443–70.

Yao, S. (1999). On the decomposition of Gini coefficients by population class and income source: A spreadsheet approach and application. *Applied Economics*, *31*(10), 1249–1264. https://doi.org/10.1080/000368499323463

Ye, P., Hu, X., Yuan, Y., & Wang, F.Y. (2017). “Population Synthesis Based on Joint Distribution Inference without Dis-aggregate Samples,” JASSS (2017). https:// doi. org/ 10. 18564/ jasss. 3533

Zacharias, A., Masterson, T., Rios-Avilla, F., Nikiforos, M., Kim, K., & Khitarshvili, T. (2019). *Macroeconomic and Microeconomic Impacts of Improving Physical and Social Infrastructure: A Macro-Micro Policy Model for Ghana and Tanzania* (September). Levy Economics Institute of Bard College. http://www.levyinstitute.org/publications/macroeconomic-and-microeconomic-impacts-of-improving-physical-and-social-infrastructure-a-macro-micro-policy-model-for-ghana-and-tanzania

Zang, E., Yang, Y. M., & Calarco, J. M. (2022). *Patterns in Receiving Informal Help with Childcare Among US Parents During the COVID-19 Pandemic* (SSRN Scholarly Paper No. 4108245). https://doi.org/10.2139/ssrn.4108245