A Modern Structural Estimation of Child Care and Women's Labor Decisions

Jesus Pacheco* Dave Foote†

March 2020

(version 20.01.a)

Abstract

This paper discusses the empirical relations of labor supply and child care utilization of mothers of young children. We based our estimations on David C. Ribar's paper 'A Structural Model of Child Care and the Labor Supply of Married Women' with updated data from the Survey of Income and Program Participation. We compared Ribar's result with our a a multinomial logit approximation to compute maximum likelihood estimators of probabilities of labor participation and child care utilization. Our estimations suggest that managing child care costs could increase mothers' labor force participation more than a comparable increase in available wages.

keywords: Labor economics, structural estimation, child care, family economics.

*University of Chicago, Harris School of Public Policy, 1307 E 60th St, Chicago, IL 60637, (312) 200-1408, jesuspachecov@uchicago.edu.

†University of Chicago, Harris School of Public Policy, 1307 E 60th St, Chicago, IL 60637, (562) 619-0580, dfoote@uchicago.edu.

1 Introduction

A family's decision making process regarding child care is a nuanced issue that can be impacted and affected by a wide range of both economic, and deeply personal considerations. The results of these decisions, however, have strong economic impacts that affect a nation at large. In this paper we apply David C. Ribar's structural model to the latest available income and child care data in order to evaluate what has and has not changed since his paper was published in 1995. Adopting a structural model for this research allows us to be explicit in our assumptions about how economic factors are guiding family decisions without making strong assumptions about the deeply personal side of these decisions. Our study aims first, at finding the main predictors of female labor participation and child care utilization acknowledging the fact that paid care utilization is an endogenous decision which is made simultaneously with other employment decision; and second, at unveiling the extent to which child care cost acts as a barrier to employment.

Our methodology for estimation consisted of solving a multinomial logit model using maximum likelihood estimation to calculate the probability of a mother to fall into each of our specifications. We used an expansive set of attributes to inform this estimation but ultimately our most effective model utilized education, median child care costs in the state, median wages in the state, the mandated staff to child ratio in the state, and binned child count columns representing very young (under 2) and other young children (3 to 5).

This estimation led us to a model that behaved quite like David C. Ribar's, wherein the cost of child care had a strong relationship with utilization of paid child care but a very weak relationship with the amount of hours a mother worked. This suggests that for the women who do work, their labor force attachment is quite inelastic. Further, child care is a normal good for these families and is purchased when it fits in the budget. This stresses the importance of quality, affordable child care.

Later in this paper, we will explain in great detail our model, our data, and finally our estimation and results.

2 Model

Previous work in labor economics, particularly on family labor supply has focused on discrete choice models (van Soest (1995); Ribar (1995)). Ribar estimated a structural discrete choice model of married mothers' care arrangements and labor supply with data of the 1980's. His estimations suggest that cost of paid care has a small effect on labor supply of mothers, but larger negative effects on paid care utilization. This model maximizes the following likelihood function of multinomial choices:

$$LF = \prod_{i=1}^{N} (\sum_{j=1}^{5} d_{ij} Prob(d_{i,j} = 1))$$

with fully specified equations for the probabilities of people in 5 different types of household in terms of their actual labor and child care decisions, that are solved with dynamic programming from the next basic setup:

$$max_{H,F} U(Y(H,F), H, F),$$

where H and F refer to the amount of hours and child care decisions respectively, and Y is income of a particular household. It is important to note that male labor decisions are exogenous factors and thus not relevant for the maximization problem.

Our model suggest a simpler interaction of factors that might have effects on labor and care allocations. We use a simplified version of this idea to estimate if these effects are detectable with more recent data. We allow for 4 possible states for each woman with young children:

- $d1$: $H = 0, F = 0$ (No labor participation, no child care utilization)
- $d2$: $H = 0, F = 1$ (No labor participation, child care utilization)
- $d3$: $H = 1, F = 0$ (Labor participation, no child care utilization)
- $d4$: $H = 1, F = 1$ (Labor participation and child care utilization)

Different to Ribar's work, we are allowed to estimate the second state ($d2$) because we have data on women using paid care even when they are not part of the labor force. This is important because our model allows estimates for complete and exhaustive subsets of the households without imposing extra estimates for sample selection.

We split our data on these four groups, and define the probability for each household with a linear logistic transformation:

$$Pr(d_i = 1|X_i, \beta) = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}},$$

where X represents a set of relevant factors and β their correspondent coefficients.

We focus our attention on two elements of X: average cost of available child care services and wage. In other words, given our assumptions, we can compare the likelihood of changing the individual's state following an increase on available wage in the job market or a decrease on the price of child care services, holding everything else constant. The rest of elements of X are relevant factors such as: age of the kids, education, hours worked and child care regulations in the state.

We characterize the general likelihood function for each state as the probability that the observed outcome came from the assumed distribution given parameter values:

$$\mathcal{L}(y_i, X_i|\beta) = \prod_{i=1}^{N} Pr(y_i = 1|X_i, \beta)^{y_i} \left[1 - Pr(y_i = 1|X_i, \beta)\right]^{1-y_i}$$

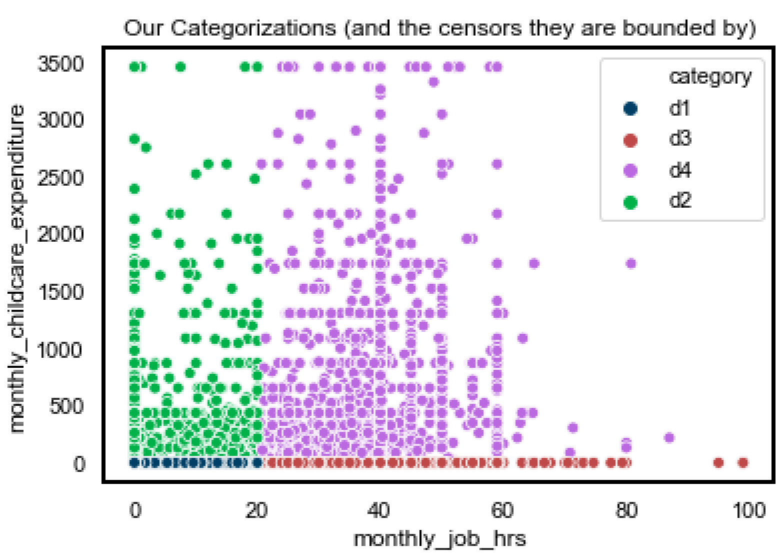
Hence, our maximum likelihood estimated β vector is defined as:

$$\hat{\beta}_{mle} = \beta : \quad \max_{\beta} \ln\left[\mathcal{L}(y_i, X_i|\beta)\right]$$

This approximation has the advantage that we can compute all the coefficients for all the possible states, but also allows for different coefficients (and different X's) for households in different states. The linearity assumption is probably the most important limitation of the model, but this project can be seen as a template, and this assumption can be gradually relaxed in future projects.

3 Data

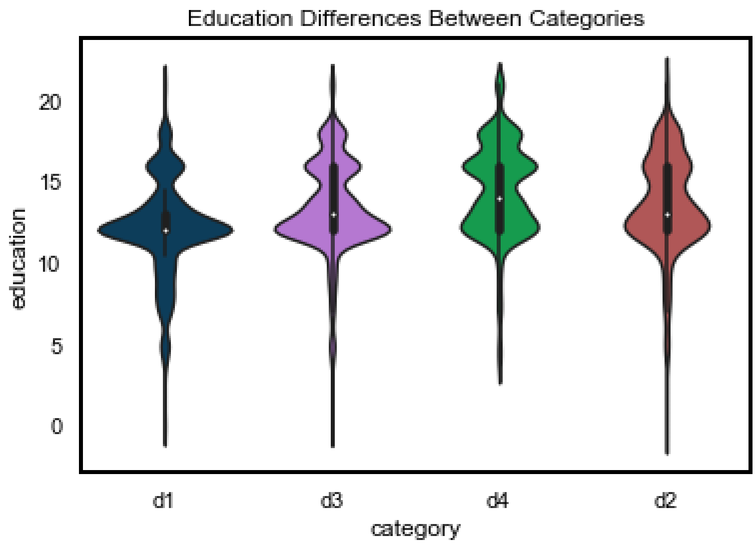
The bulk of our variables were derived from the Survey of Income and Program Participation, which is openly available in csv, sas7bdat, and dta format on the National Bureau of Economic Research website. The individual surveys are conducted over a period of two and a half to four years and are composed of continuous waves of 14,000 to 52,000 households determined to be 'representative' of the United States. All members of a participating household aged 15 or older were interviewed by self-response when possible, although some individuals were interviewed 'by-proxy' - meaning another household member answered for them. For this study, we have utilized the 2014 Wave 1 Core data, available through the National Bureau of Economic Research. The survey was especially useful for our study as it collects rich and specific data on income (including wages, investment vehicles, insurance payouts) and also tracks childcare expenditure. This and all other SIPP data are created and released by the United States Census Bureau, and represent the anonymized results of the survey. In order to achieve anonymity in a data set with such a great level of detail, the Census Bureau had to top-code assets, earnings, and medical expenses. The median, mean, and standard deviation of those values which do fall outside of the uncensored range are available, but for our study recreating and randomly distributing these values threatened to introduce unknown and possibly hallucinatory biases.



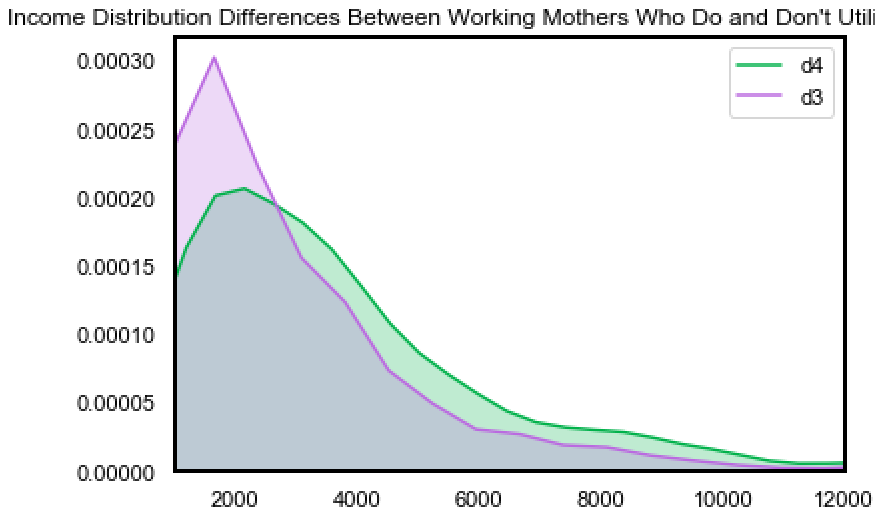
The SIPP has been used for a wide array of studies: from comparing the wealth holdings of US and foreign born citizens (Clark-Hildebrand), to helping understand retirement plan participation (Copeland), and of course a wide array of women's labor supply and childcare studies, such as the David C. Ribar paper "Child Care and the Labor Supply of Married Women", the methods from which we are replicating in this paper. This survey continues to be a useful tool for research because of its high level of detail. The released data set contains 3,133 variables gathered for each of the 15 and older members of a responding household, and offers a large degree of granularity on income, employment, government and employer program participation for respondents and the people they live with.

Exact replication of our extraction and transformation of this data set can be achieved by following the instructions on our Git repository found here. As you can see, we used both Python and R to achieve the final dataset we wanted. Due to the large amount of data wrangling required to use the SIPP data set, we think it is important that we explain these transformations explicitly for the portion of our audience not fluent in both languages. The SIPP metadata description, featuring the most accurate possible descriptions of their variables, can be found online as well.

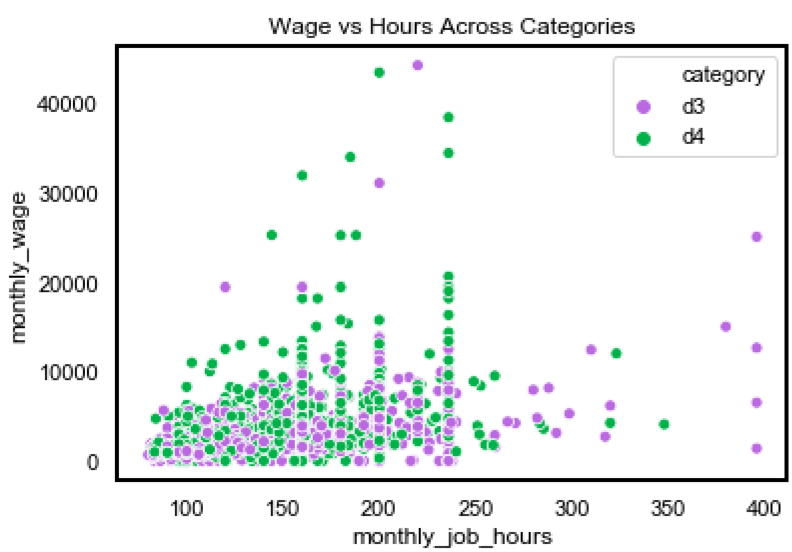
Our education variable, discoverable as 'EEDUC' in the SIPP Wave 1 Core CSV file, was the first choice we had to make in our encoding of SIPP data for this analysis. We chose to represent education bands in our data with the minimum amount of education required to appear in that category, measured in grades/years beyond kindergarten, with the exception of people whose education ended in the 1st-4th grade range, to whom 4 years of education were credited. This means women who completed high school with no college experience were binned together with women who completed high school with less than a year of college experience, but both are differentiated from women with at least a year of college experience. As is custom in the United States, we credited 2 years for an Associate's degree, 4 years for a Bachelor's, an extra 2 years on top of that for a Master's or Professional School degree (e.g. Medical, Public Policy, and Law Schools), and an additional 5 years beyond a Bachelor's degree for all Doctorate degrees. We acknowledge five years for a Doctorate level degree is somewhat arbitrary but the figure simply allows us to categorize respondents who have attained Doctorates without an in depth study into the mean years to completion across all current fields and programs.



Our income variables were also derived from the SIPP data set, which categorized types of income at a level of granularity beyond what was needed for our study. To achieve the categorization we were looking for, earned vs unearned income, we aggregated wages, disability payments earned through any kind of employment activity, and non-wage income derived from work. Our rationale for including employment based disability payments was that these payments are payed in situations where the recipient is not likely to be capable of all child care responsibilities that may come up, and thus still represent the value of a trade off between work and family duties. Social security payments are also generally directed to people with young children when they are no longer able to work, but have worked, and are in a circumstance where their ability to care for a child is likely to be limited as well (old or sick). Our guiding principle in this classification was to mark all income earned by a situation that limited a respondent's ability to provide complete childcare as earned income. We then calculated unearned income as the difference between total personal income (TPTOTINC in the data) and our figure for earned income.



Other variables we generated from the SIPP 2014 Wave 1 CSV included monthly childcare expenditure, monthly binned number of children metrics and monthly work hours. Number of children under two, and from 3 to 5, were calculated from the 20 'TCBYR' columns in the original data set, which represent the birthyears of up to 20 children for a given respondant (although the most children someone had in our data was 8). Monthly job hours map through a x4 transformation to the data set's 'TMWKHRS' variable, which represents hours that the respondent worked in a week.



We used two additional sources to generate variables for our study. To account for a measurement of the state regulations for child care services, we used the aggregate information available in the U.S. Department of Health and Human Services - Office of Child Care Services (2014). Since most of the regulations are somewhat homogeneous in most states, we used the child:staff ratio where we find some variation in that variable. The regulations on the amount of staff required per every child in a child care facility ranges from 4 (tightest regulation) to 12 (loosest regulation). The data on median wages are retrieved from the U.S. Bureau of Labor Statistics (2018) ant they are aggregated by state. These variables were added to the data with our R script, which is also available on our Github page.

4 Results and analysis

Table 1 presents summary statistics of our variables of interest for mothers with children 15 years old and younger in 2014. We can see that half of the women in our sample are working outside of the household, and about a third are using at least some paid care. The maximum likelihood estimation results are shown in Table 2.

Variable	**Mean**	**Std.**
Working	0.510	0.499
Utilizing child care	0.314	0.464
Kids under 2	0.341	0.527
Kids 3-5	0.470	0.599
Education	13.180	2.993
Staff:Child ratio	8.298	2.033
Workhours per month	21.373	18.913
Median childcare cost (in state)	10.554	1.289
Median wage (in state)	18.13	1.957
Number of observations	5,908	

Table 1: Summary Statistics

Model predictions. From our estimation, we can see that children's age are the major predictors in this model. For household with working mothers ($d3$ and

$d4$), having young children make them more likely to use paid care. Education, as expected, reduces the probability of not working and not using paid care as the opportunity cost of leisure/household work becomes higher. Furthermore, the biggest effect of increasing childcare costs will be in the increase of women working and not using paid child care; while the biggest effect of increasing wages will be in the increase of women not working but being able to afford paid child care, which we interpret as a potential income effect through spouses income (exogenous to our model). It is important to point out that our point estimates are robust to different specifications, but standard errors were not. The standard errors presented in Table 2 where computed using the GLM-Logit function in *R* for that reason.

Increasing labor participation. The main result of our estimation is that managing child care costs could increase mothers' labor force participation more than a comparable increase in available wages. While a perfect representation of these mechanisms is somewhat beyond our model, our specification allows us to differentiate between these two considerations mothers make when maximizing the utility of their families. This differentiation highlighted the fact that women's labor force participation is more strongly related to the cost and availability of child care than their potential wages. Further, for mothers who are already participating in the labor force, high child care costs will reduce utilization to a greater extent than it will discourage labor participation. The implication here is that pricing working mothers out of the child care market will directly cause children to spend more time unsupervised and alone. The effects are somewhat small compared to other more personal determinants of labor decisions, these are the areas with labor policy implications. Our results are in line with Ribar's result from the 80's in the idea that an increase of child care costs seem to be associated with a greater decrease on utilization than in labor participation. We acknowledge the limitation of measuring child care cost as the median wage of workers in the child care occupation, but given the lack of a better measure, we believe this interpretation could be legitimate and fruitful for the analysis, with the caveat that by reduction of child care costs we do not mean

Variable	d1 $H=0, F=0$	d2 $H=0, F=1$	d3 $H=1, F=0$	d4 $H=1, F=1$
intercept	6.230 (0.773)	-4.544 (0.750)	-1.697 (0.559)	-5.525 (0.585)
kids under 2	0.213 (0.092)	-0.192 (0.095)	-0.273 (0.073)	0.260 (0.074)
kids 3 to 5	-0.466 (0.079)	0.487 (0.075)	-0.562 (0.065)	0.574 (0.063)
education	-0.136 (0.018)	0.158 (0.018)	-0.127 (0.013)	0.096 (0.014)
staff:child ratio	-0.046 (0.030)	-0.003 (0.031)	0.043 (0.022)	-0.023 (0.022)
monthly working hours	-0.044 (0.004)	-0.017 (0.004)	0.023 (0.003)	0.022 (0.003)
median child care cost (in state)	-0.063 (0.086)	-0.054 (0.087)	**0.091** (0.062)	-0.057 (0.065)
median wage (in state)	-0.051 (0.058)	**0.073** (0.057)	-0.060 (0.041)	0.050 (0.043)
Log likelihood	1,434.12	5,177.39	9,976.21	1,0474.83
No. of observations	2,399	493	1,655	1,361

Table 2: Estimation results

reducing the wage of child care workers. Another improvement for further studies could be finding better aggregations of the mean wages as we understand that the state-level data is quite heterogeneous.

5 Conclusion

We estimated a logistic model of labor participation and child care utilization using data from the Survey of Income and Program Participation 2014. The results were that decreasing child care costs might have a greater effect towards labor participation than increasing available wages for mothers of young children. In any case, lower costs of child care will increase more utilization than labor participation, confirming

Ribar's claim in 1995 with our more sophisticated discrete choice model that accounts for non-working mothers who pay for child care.

Child Care Aware of America (2018), a group that advocates on behalf of parents who require child care, estimated that 11% of married and 37% of single parents' incomes were committed to child care expenses in 2018 with an average cost between \$9,000 and \$9,600. This represents a significant burden for working families, and due to the fact that labor is a large part of the cost of child care (and a huge determinant of the quality of child care), increases in wages can fail to translate to better child care utilization among working women. In our research, we noticed this play out between groups d3 and d4, our working mothers who did and did not utilize child care, respectively. The median cost of childcare, more than the median wage in the state or even the hours a mother worked, was the best predictor of whether or not a woman might utilize child care. This fact underscore the importance of finding ways to provide more affordable child care services across the country but especially in places where child care is most expensive. Many federal aid bills in the United States target low incomes, but in this case it is more prudent to find safe ways to change the cost structure for providers, without touching wages.

Our findings also stress the potential of government initiatives like Head Start and after school programs to increase labor participation and in effect household incomes in a multiplicative way. Providing free childcare services allows mothers who do not work to enter the labor force while also improving outcomes for younger children currently spending part of their days unsupervised. Currently only three states have any regulations on the books concerning restrictions to leaving children unsupervised, and according to Hartwell-Walker (2018) in her article "Children Who are Home Alone", 7 million of the nation's 38 million children ages 5 to 14 are left home alone regularly. Reducing this figure offers working parents more freedom to concentrate on their work without worrying about their children, while also improving educational and health outcomes for children as they grow. Unfortunately, as the cost of child care continues to rise we expect to see a growing number of children left unsupervised for long periods of time. Part of the future research we would like to see outside of the realm of labor economics is an inquiry into the effect of time spent alone on these children.

Future research should attempt at relaxing the linearity assumptions of our model with fully specified model solutions as the discrete choice model one. One feature we believe would enrich the discussion is to incorporate complementarities to a child care production function between paid child care utilization and time of parenting at home (not working). Further simulations can be made to challenge this model and incorporate further waves of the SIPP in the data.

References

Child Care Aware of America, "The US and the High Cost of Child Care," 2018. https://cdn2.hubspot.net/hubfs/3957809/COCreport2018_1.pdf.

Hartwell-Walker, Marie, "Children Who are Home Alone," *Psych Central*, 2018. https://psychcentral.com/lib/children-who-are-home-alone/?all=1.

Ribar, David C, "A structural model of child care and the labor supply of married women," *Journal of Labor Economics*, 1995, *13* (3), 558–597.

U.S. Bureau of Labor Statistics, "Wage Data by Area and Occupation," Apr 2018. https://www.bls.gov/bls/blswage.htm.

U.S. Department of Health and Human Services - Office of Child Care Services, "Data Explorer and State Profiles," 2014. https://childcareta.acf.hhs.gov/data.

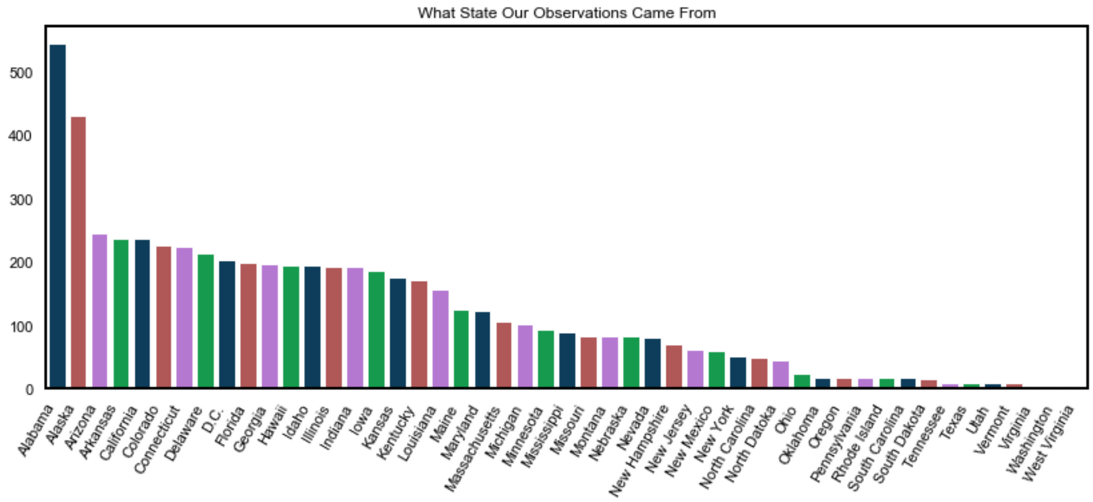
van Soest, Arthur, "Structural Models of Family Labor Supply: A Discrete Choice Approach," *The Journal of Human Resources*, 1995, *30* (1), 63–88.

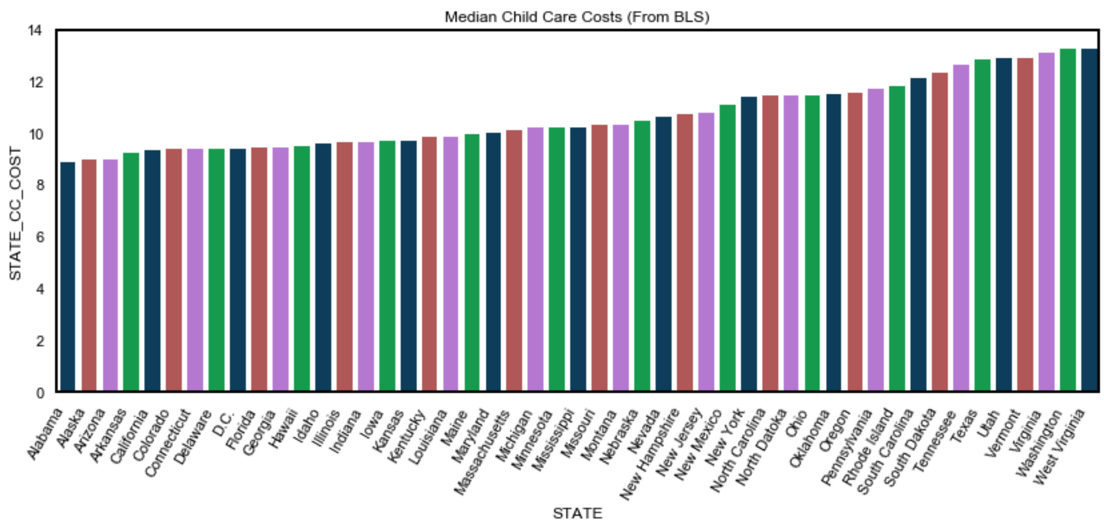
A Technical Appendix

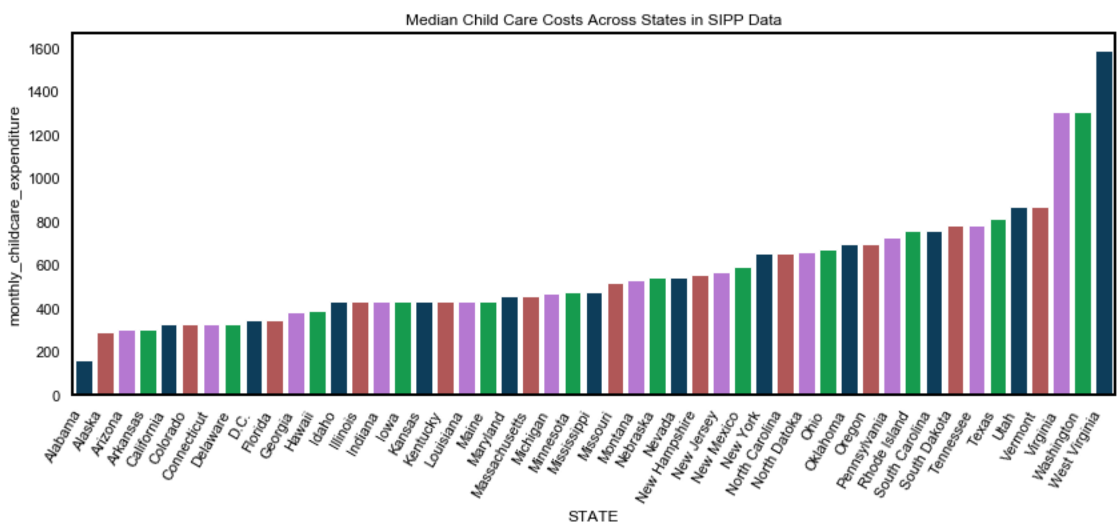
Differences in New SIPP Data Releases Key differences between our estimation and the original David C. Ribar structural model are centered around our use of the newest Survey of Income and Program Participation's Wave 1 data, which was gathered in 2014. Our predecessor used the 1986 iteration of SIPP, which had some minor differences. The most important of these differences was the change from reporting hours of child care utilization to reporting monthly expenditures on child care, which limited the data we had on free child care arrangements. This causes us little concern, however, since just like our predecessor we treated circumstances preventing the need for paid child care as exogenous and the decision to purchase child care as an endogenous choice married women make in the maximization of their family's utility.

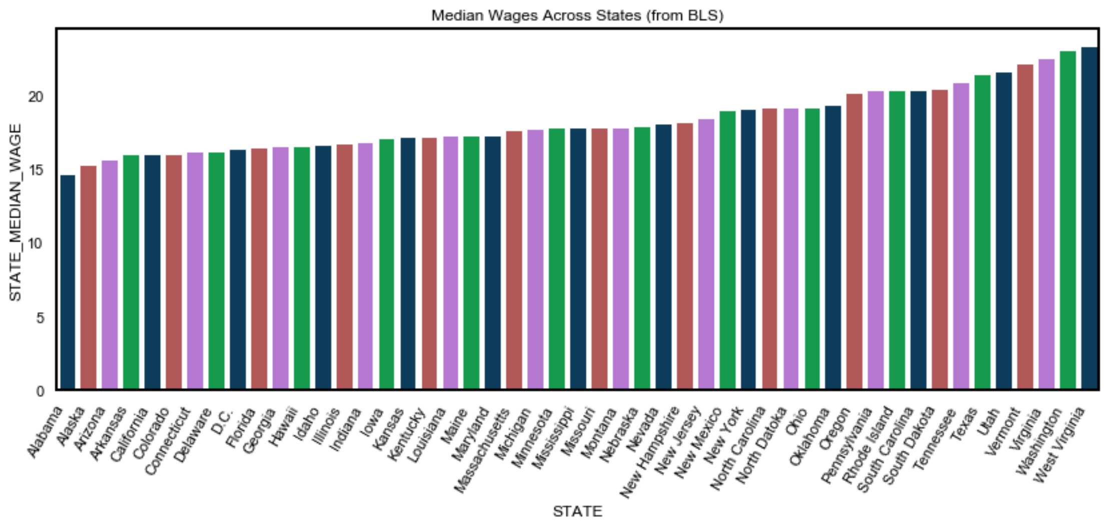
The other key difference was our decision to consider 4 possible states of child care and employment for each woman. These states were labelled $d1$, which represented women who did not participate in the labor force to any significant degree and also did not use paid child care; $d2$, for women who did not participate in the labor force but did utilize child care (left out from the previous study but very present in our data set); $d3$ for working mothers who did not pay for child care; and $d4$ for the working mothers who did pay for child care. In the earlier paper, part-time employment was classified separately but mothers who did not work while paying for child care were not considered. This simplification allowed us to include all mothers in from our data set in this estimation.

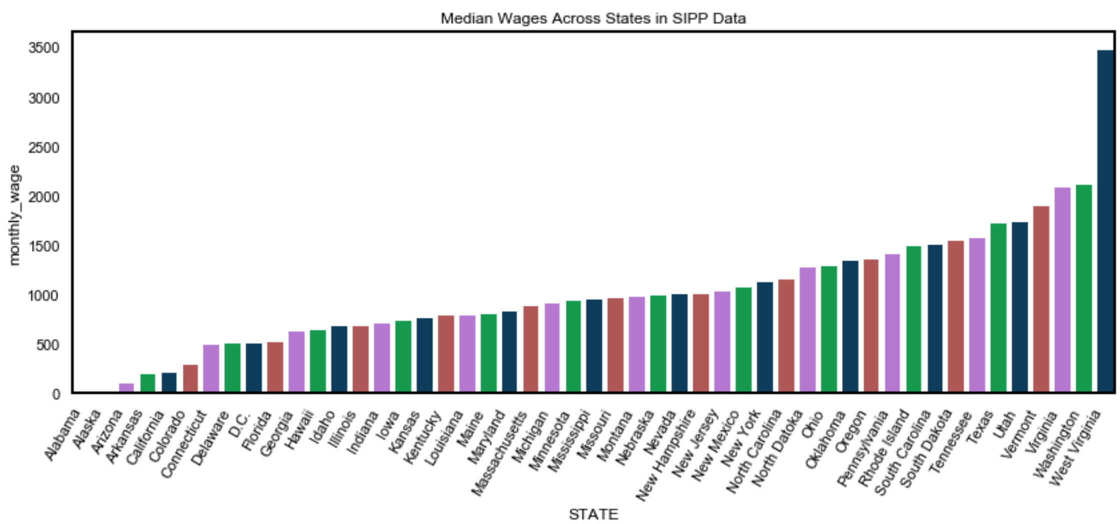
Visual Gallery

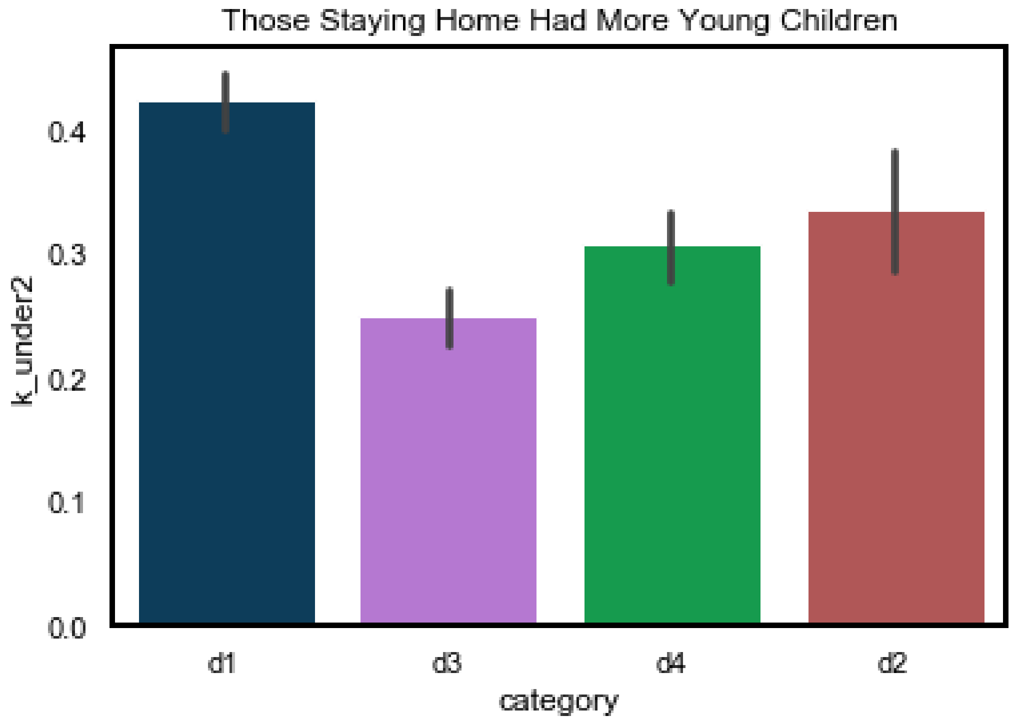


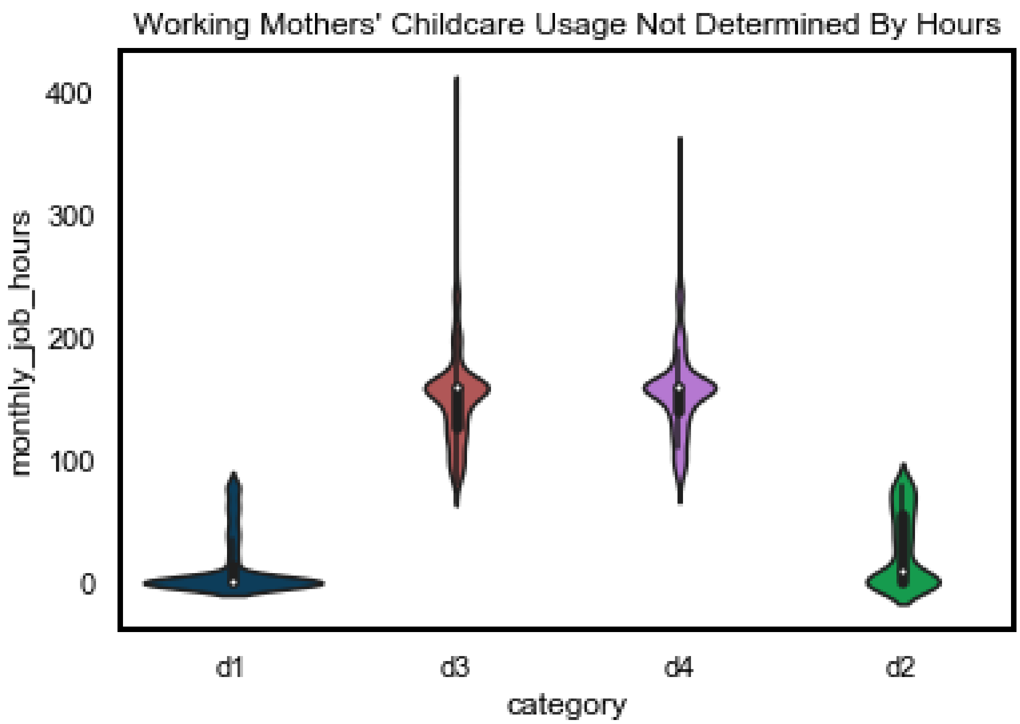


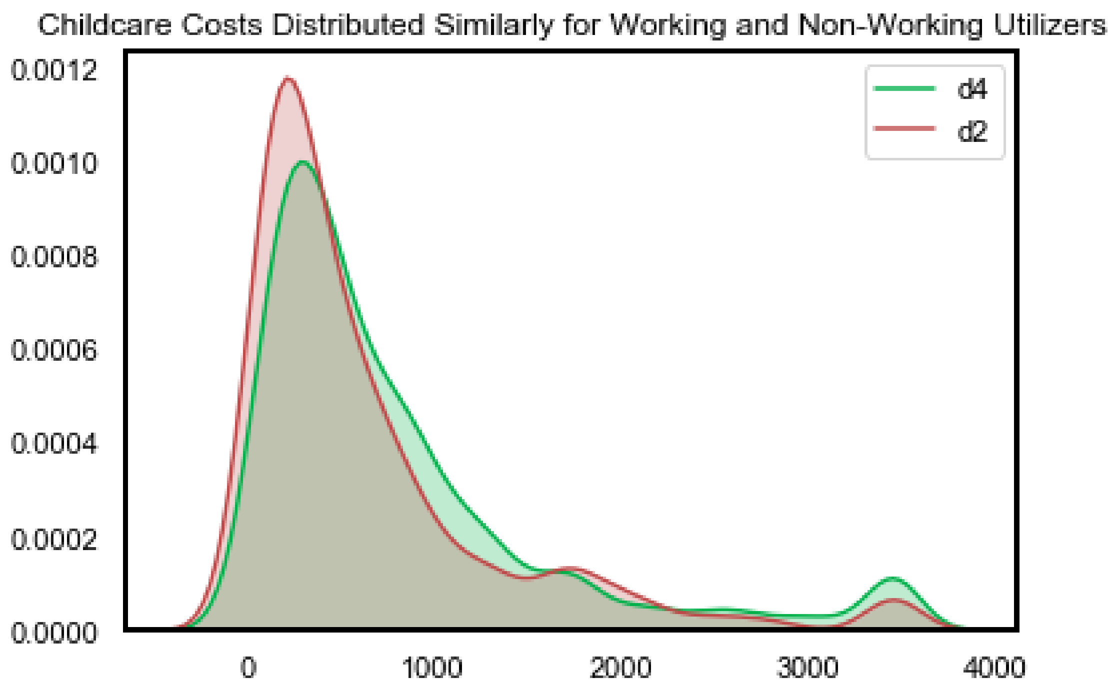



Median Wages Across States (from BLS)
STATE_MEDIAN_WAGE
0
5
10
15
20
STATE
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
D.C.
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Datoka
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia


Median Wages Across States in SIPP Data
monthly_wage
0
500
1000
1500
2000
2500
3000
3500
STATE
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
D.C.
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Datoka
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia


Those Staying Home Had More Young Children
0.4
0.3
0.2
0.1
0.0
k_under2
d1
d3
d4
d2
category


Working Mothers' Childcare Usage Not Determined By Hours
400
300
200
100
0
monthly_job_hours
d1
d3
d4
d2
category


Childcare Costs Distributed Similarly for Working and Non-Working Utilizers
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
1000
2000
3000
4000
d4
d2


Having Children Under 2 Affects Working Women Differently
monthly_childcare_expenditure
4000
3000
2000
1000
0
False
True
has_kunder2
category
d4
d2