Labor Supply Heterogeneity and Demand for Child Care of Mothers with Young Children

Patricia Apps
Sydney University Law School and IZA

Jan Kabátek*
Tilburg University and Netspar

Ray Rees
CES, University of Munich

Arthur van Soest
Tilburg University and Netspar

February 23, 2012

Abstract

This paper introduces a static structural model of hours of market labor supply, time spent on child care and other domestic work, and bought in child care for married or cohabiting mothers with pre-school age children. The father's behavior is taken as given. The main goal is to analyze the sensitivity of hours of market work, parental child care, other household production and formal child care to the wage rate, the price of child care, taxes, benefits and child care subsidies. To account for the non-convex nature of the budget sets and, possibly, the household technology, a discrete choice model is used. The model is estimated using the HILDA dataset, a rich household survey of the Australian population, which contains detailed information on time use, child care demands and the corresponding prices. Simulations based on the estimates show that the time allocations of women with pre-school children are highly sensitive to changes in wages and the costs of child care. A policy simulation suggests that labor force participation and hours of paid work would increase substantially in a fiscal system based solely on individual rather than joint taxation.

JEL classification: J22, J13, H24
Keywords: Time use, income tax, child care subsidies

*Corresponding author: Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.kabatek@uvt.nl.

Preliminary version

1 Introduction

One of the most striking, and still largely unexplained, facts about female labor supply in the developed countries is its heterogeneity across households, and indeed across countries. In most OECD countries, on average around one-third of women work full time in the labor force, one third do various amounts of part time work, and one third work solely in household production. Very little of the aggregate heterogeneity across all households in any one country is explained by wage rate differences and by the number of children present in the household. Moreover, the correlation between female labor supply and fertility across these countries is strongly positive, even though historically, in any one country, there has been an inverse relationship between them.

Some insight is gained by organizing the data in terms of life cycle phases based on the number and age of children in the household. In the pre-children phase, there is very little difference between male and female labor supply distributions. This changes dramatically when children enter the household, and this is when the female labor supply heterogeneity essentially sets in. Though there is a trend of return to the labor force over subsequent phases of the life cycle as the children reach school age and beyond, the basic pattern of heterogeneity persists. Such findings suggest that for the theoretical and empirical analysis of the female labor supply heterogeneity it would be fruitful to focus on the life cycle phase in which the household makes its labor supply decisions in the presence of young children. However, in order to do so, it is important to be aware of idiosyncratic characteristics which are common among these households.

Clearly, the birth of a child represents a fundamental change in the lives of the parents. Childbearing affects the couple's preferences, consumption patterns, and it is particularly demanding in terms of their time. The neccessity to nurture and educate the young child leads parents either to specialize in self-provision of childcare, or to rely on services provided by other caregivers such as relatives, friends, or formal institutions.

Sufficient provision of childcare is essential for proper development of the children, and as such, it is bound to have profound impact on the mother's decision making, especially in terms of labor force participation. The importance of the availability and the cost of childcare services has been confirmed by a host of both theoretical (Apps & Rees 2009) and empirical studies[1] of the female labor supply. And for that reason, we consider careful treatment of the mothers' childcare-related decisions to be one of the key aspects of female labor supply modeling.

This paper presents a static structural discrete choice model analyzing the time allocation choices of married and cohabiting mothers with pre-school aged

[1] Previous works on labor supply and childcare decisions include Ribar (1995), Duncan *et al.* (2001), Blau (2003), Connelly & Kimmel (2003), Doiron & Kalb (2005), Kalenkoski *et al.* (2005), Kornstad & Thoresen (2007), Baker *et al.* (2008), and Blundell & Shepard (2011).

children[2]. The main advantage of the discrete choice approach is that it can account for the non-convex nature of the household budget sets and, possibly, also of the household technology. Within the model, we analyze mother's decisions about her hours of market work, time spent on childcare and other domestic work, and amount of bought-in childcare. The main goal is to assess the sensitivity of choices at the intensive and extensive margin of the female labor supply, and to capture underlying substitution patterns between alternative uses of mothers' time.

Unlike previous studies, we allow the household utility to take on more flexible functional form (following Van Soest (1995) and Van Soest & Stancanelli (2010)). We also embody both formal and informal childcare directly into the household utility function[3]. The key innovation of our approach is however the incorporation of unobserved heterogeneity in the flexible form of latent classes[4]. We choose to extend the treatment of unobserved heterogeneity beyond the traditional framework of random coefficient models[5], as the assumptions imposed on the stochastic parameter components may well prove unrealistic in the context of female labor supply.

The reliability of random coefficient models hinges on the assumption that the latent heterogeneity is sufficiently well-behaved. That way, the practitioner can infer its approximate distribution and dimensionality, and subsequently allow the appropriate coefficients to take on more flexible stochastic form. However, this approach often fails to improve the estimation results, with an example being Doiron & Kalb (2005), who conclude that in the context of Australian female labor supply it is unnecessary to allow for unobserved heterogeneity[6]. We aim to explore the effects of unobserved heterogeneity without relying on strict distributional assumptions of the random coefficient model, and assess its importance by comparing relative performance of the model with and without latent class heterogeneity.

The model is estimated using the HILDA dataset, a household survey of the Australian population which contains detailed information on time use, childcare demands and the corresponding prices. Simulations based on our estimates show

---

[2]Similar modelling framework is also employed by Doiron & Kalb (2005), Kornstad & Thoresen (2007), and Blundell & Shepard (2011)

[3]In prior works, the utility function is either allowed to accomodate only formal childcare (Ribar 1995), or it restricts the childcare to affect utility indirectly (Doiron & Kalb (2005), Kornstad & Thoresen (2007)). The indirect effect is achieved through the cost channel, as utilization of bought-in care lowers the disposable household income. Our approach instead asserts that the utility derived from formal childcare goes beyond its monetary costs, and emphasizes its dependence on the availability of other, informally provided childcare.

[4]As in Train (2008) and Pacifico (2009).

[5]Applications using this approach include Ribar (1995), Doiron & Kalb (2005) or Van Soest & Stancanelli (2010)

[6]There are two interpretations of such finding: either the model succeeds in capturing most of the heterogeneity, rendering the stochastic terms redundant, or the model imposes wrong distributional assumptions on the unobserved heterogeneity and fails to capture its true effect. For that reason, we believe that instead of rejecting the concept of unobserved heterogeneity *per se*, one should firstly examine more general modelling approaches, controlling for potential misspecification of the original model.

that hours of market work and the formal childcare demands of mothers with pre-school children are highly sensitive to changes in net wages and the costs of bought-in childcare. A policy simulation suggests that labor force participation and hours of market work would increase substantially in a fiscal system based solely on individual rather than joint taxation.

The paper is organized as follows. In the next section we set out the underlying household model. In Section 3 we present the econometric specification of the model that we take to the data. Section 4 discusses the data used for the estimations and Section 5 presents and discusses the estimation results. Section 6 presents the results of the simulations. Section 7 concludes.

2 The Model

We present the model of household choice in a static framework, ignoring the fact that this is just one phase in the household's overall life cycle. This seems a strong limitation, since *a priori* we expect that decisions made in this phase, especially on female labor supply, could be influenced by intertemporal factors, such as the anticipated loss of human capital resulting from reducing current market labor supply and the effects of this on future wage rates and employment possibilities. Thus a woman may continue working in this phase, despite the low or even negative current wage net of tax, social security payments and childcare costs, as an investment in her long-term career prospects. Since lack of data precludes incorporation of these issues in the econometric work, we cannot capture them fully in the following model. However, we do attempt to capture them in a reduced form sense, since the marginal utility of market work *vis á vis* leisure or domestic work will tend to be increased by the existence of such factors. We also take the number of children in the household as exogenously fixed, and so do not model the fertility decision.

Household $h = 1, 2, ...H$, chooses:

- its consumptions x_{ih}, with $i = 1, 2, .., n$ denoting the individuals within the household;
- the mother's leisure consumption l_{2h};
- total childcare z_h, comprising both formal and informal childcare for all children;
- consumptions of a non-childcare household good y_{ih};
- the second earner's time inputs into childcare and production of the household good t^c_{2h} and t^y_{2h} respectively;
- purchases of the market childcare good m^c_h.

Consumption is a composite market good with price 1, the mother's gross wage rate is w_{2h}, and the price of the market childcare good is p_h^c. Note that, as the data suggest, we allow this price to vary across households.[7] Throughout, we take the father's leisure $\hat{l}_{1h}$ and time allocations $\hat{t}_{1h}^c$, $\hat{t}_{1h}^y$, as exogenously given. Therefore, given the time endowment constraint, his market labor supply $L_{1h} = \hat{L}_{1h}$ is also taken as exogenously fixed. The sum of the primary and second earners' gross incomes from market supply, $\sum_i w_{ih}L_{ih}$ is denoted by $I_h(w_{1h}, w_{2h})$. The two adults in the household have utility functions $u_{ih}(x_{ih}, y_{ih}, l_{ih})$, $i = 1, 2$, and the childrens' utilities are $u_{ih}(x_{ih}, y_{ih}, z_h)$, $i = 3, .., n$. Thus childcare is modeled as a household public good that determines the utility level of the children.

The household is assumed to maximize a household welfare function, concave in utilities,

$$W_h = \Psi_h(u_{1h}(.), ..., u_{nh}(.); \mathbf{e}_h) \;\; h = 1, 2, ...H \tag{1}$$

where $\mathbf{e}_h$ is a vector of exogenously given "environmental" or "distributional" factors which can be interpreted as determining the household's preferences over the utility profiles of its members.[8] This function is based upon some household choice process which need not be further specified, and is intended to capture such things as love and caring for each other, as well as more conventional attributes of social welfare functions such as ethical views of fairness.

The household's budget constraint can be written as

$$\sum_i x_{ih} + p_h^c m_h^c \leq I_h(w_{1h}, w_{2h}) - T(I_h(w_{1h}, w_{2h}), p_h^c m_h^c; n, ..) \;\; h = 1, 2, ...H \tag{2}$$

where $T(.)$ is a tax/benefit function which may contain as arguments demographic variables as well as gross incomes and expenditure on bought in childcare.[9]

The technology of household production is expressed by the production functions

$$z_h = f_h(\hat{t}_{1h}^c, t_{2h}^c, m_h^c) \;\; h = 1, 2, ...H \tag{3}$$

$$\sum_i y_{ih} = y_h = g_h(\hat{t}_{1h}^y, t_{2h}^y) \;\; h = 1, 2, ...H \tag{4}$$

and there is a time constraint

$$l_{2h} + t_{2h}^c + t_{2h}^y + L_{2h} = T \tag{5}$$

where T is a given time endowment. The main restrictive assumption on the production functions is the separability across outputs z_h, y_h, which rules out joint production. Because we will be adopting a discrete optimization approach,

[7] Every variable or function with subscript h can vary across households. Each of these is therefore in principle a contributor to across-household heterogeneity in choices.

[8] In principle, the distributional factors could also include the wage rates, but this will not be allowed for in the empirical model.

[9] For example there may be tax offsets for expenditure on market child care.

directly comparing values of the household welfare function at all choice opportunities (see Van Soest, 1995), we do not need to impose conditions of convexity or even differentiability on the various functions in (2), (3), and (4). Thus the household can be thought of as choosing the variables m_h^c, l_{2h}, t_{2h}^c and t_{2h}^y that determine consumptions, market labor supplies and income *via* the constraints (2)-(5) in such a way as to yield a global maximum of the function $\Psi_h(.)$.

3 Econometric Model Specification

We base the econometric specification[10] on three choice variables: hours of mother's market work, hours of mother's housework and hours of bought-in childcare. In order to employ discrete choice methods, we restrict these variables to take one of five possible numerical values, which we can characterize as "low", "low-medium", "medium", "high-medium" and "high". This yields a grid of $5^3 = 125$ possible discrete choice points. For the purpose of our model we specify the vector[11] $\mu = [l_2, t_2^y, m^c, Y]$, with the leisure variable, l_2, derived as the residual of the daily time constraint (24 hours) after subtracting the market work and household hours.

The fourth variable, Y, is net household income, calculated as gross income net of taxes, family tax benefits and expenditure on childcare. Gross income is the sum of each partner's earnings and the family's non-labor income. The husband's earnings and non-labor income are treated as exogenous. The mother's gross earnings are calculated as the product of a predicted gross wage rate (using a Heckman selection model; see Heckman, 1979) and her choice of market hours. Expenditure on childcare is calculated as the product of a predicted childcare price and the household's choice of childcare hours. Since household income does not include the value of household production it does not depend on the time t_2^y spent on household production. There are therefore 25 possible values of net household income for each household, corresponding to all combinations of five choices of L_{2h} and five choices of m_h^c.

3.1 Basic Model

We first present the model without unobserved heterogeneity. We take a reduced form of the household welfare function introduced in the previous section, specified as a flexible quadratic function

$$\Psi(\mu) = \mu' \mathbf{A} \mu + \mathbf{b}' \mu \tag{6}$$

where $\mathbf{A}$ is a symmetric 4×4 coefficient matrix, and $\mathbf{b}$ is a 4-component vector. The first three components of $\mathbf{b}$, corresponding to the time use variables

[10] For detailed discussion and applications of the discrete approach adopted here see, for example, van Soest (1995), van Soest, Das and Gong (2002) and Pacifico (2009)

[11] Since this formulation applies to each individual household we drop the household subscript for convenience.

l_2, t_2^y, m^c, are defined as

$$b_j = \sum_{k=1}^{K} \beta_{kj} X_k \quad j = 1,..,3 \tag{7}$$

where the X_k denote respectively a constant term and variables representing observed household characteristics such as wife's age; wife's age squared; number of pre-school age children; number of school-age children; and hours of informal childcare provided by relatives, friends or the husband. These represent sources of observed heterogeneity. The elements of the matrix $\mathbf{A}$ as well as the component b_4 are assumed the same for all households[12].

This household welfare function is in reduced form since it does not explicitly incorporate household production, the utility functions of the household members, or the household process which combines the utilities of the members. This should be kept in mind when interpreting the parameters. For example, the partial derivative of $\Psi(.)$ with respect to l_2 is the marginal change in household welfare when the other components of μ - t_2^y, m^c and Y - are held constant, that is, when an hour of market work is replaced by an hour of domestic work without changing income. This captures the (positive) effect of additional home production as well as the potential (positive or negative) effect of a higher or lower preference for domestic rather than market work, not accounting for the value of home production or the wage for market work. Differences in b_1 across households may therefore either reflect differences in productivity in household production or differences in preferences, or both. Conceptually, these are of course two quite distinct sources of heterogeneity, but they cannot be separately identified in the available data (since we do not observe the output of household production).

We introduce randomness in the value of the household welfare function at each possible choice point (l_2, t_2^y, m^c, Y) by specifying:

$$\Psi_r = \Psi(.) + \varepsilon_r \; r = 1, 2, .., 125 \tag{8}$$

We can rationalize these errors as being due to optimization errors or to unobserved alternative specific characteristics that make each alternative more or less attractive than predicted by the systematic part. They can be due to factors that make a specific alternative more (less) attractive because of high (low) productivity or other, possibly preference-related, factors. The ε_r are assumed to be independent of each other and identically distributed and to follow the Type 1 Extreme Value Distribution. This implies that the conditional probability that point r^* is chosen as the optimal point is

$$P[\Psi_{r^*} > \Psi_r \forall r \neq r^* \mid \mu, \mathbf{A}, \mathbf{b}] = \frac{\exp \Psi(\mu_{r^*}, \mathbf{A}, \mathbf{b})}{\sum_{r=1}^{125} \exp \Psi(\mu_r, \mathbf{A}, \mathbf{b})} \tag{9}$$

[12]The linear coefficient of b_4 is left without any interactions to reduce the computational complexity of the problem. We do not regard this adjustment to be very restrictive, given that the utility function is identified up to a monotonic transformation only.

Finally, to guarantee that household welfare always increases with household income (an assumption which is needed for economic interpretation of the estimates) we penalize the likelihood when necessary by adding points inside the budget frontier as additional choices (that are never chosen by the household).

3.2 Unobserved heterogeneity

It is likely that different households within the selected sample of families with young children have different unobserved attributes, for example in human and physical capital. This will lead to differences in productivity in domestic production, which may influence their labor supply choices. Incorporating this kind of heterogeneity is crucial for understanding the implications of the model. For example, heterogeneity in domestic productivity can drive the relationship between observed household income (widely used as the tax base) and household utility possibilities. A low domestic productivity household may decide to outsource home production or childcare and spend more time on market work. This household will have higher income than an otherwise similar household with high domestic productivity which does specializes in home production. Ignoring the difference in domestic productivity and the contribution of home production to household welfare would lead to the incorrect conclusion that the first household is necessarily better off.

A common approach in studies of female labor supply is to ignore household production, including childcare, entirely and to attribute heterogeneity to differences in preferences for leisure. This is problematic if the results of the empirical analysis are to be used to evaluate the policy consequences for social welfare, since preference heterogeneity constitutes a controversial and rather dubious rationale for income redistribution. Why should income be redistributed to households which have lower income levels simply because their second earner, at least, has a higher preference for leisure? Heterogeneity caused by productivity differences is not subject to this problem, because differences in productivity create real differences in utility possibilities across households, and therefore constitute a justifiable basis for income redistribution.

In any case, incorporating heterogeneity driving labor supply choices, whether in productivities or in preferences, is central to the analysis of tax reform and transfer policy, and the argument in the foregoing paragraph provides the rationale for stressing the possibility of variation in productivities, as well as prices. Since we do not have data on domestic productivities, this implies the need to deal with the problem of unobserved heterogeneity in the present study.

It is therefore important that the heterogeneity driven by unobserved variation in household productivities is captured in the specification of the model. The errors ε_r in the basic model only do this to a very limited extent, since they are alternative-specific and imply independence of irrelevant alternatives. A high productivity in domestic work will probably make all choices with high value of l_2 more attractive, something that cannot be captured by errors that are independent across alternatives.

Several alternative approaches have been developed to allow for unobserved heterogeneity in the context of discrete choice labor supply models, with the most prominent method being the parametric random coefficients model (see Van Soest, 1995, or Keane & Moffitt, 1998). However, this model is criticized for often non-justifiable assumptions imposed on the distribution of stochastic terms (see Burda et al., 2008, Train, 2008, or Pacifico, 2009). The distributions are predominantly assumed to be multivariate normal or log-normal, which implies the corresponding density of parameter values to be unimodal, that is, having one peak characterizing the most frequent household welfare function.

The unimodal approximation of the parameter distribution can however prove invalid in empirical applications. In particular, previous empirical and theoretical works (Apps & Rees, 2009) suggest that multimodal parameter distributions might well be present in the context of female labor supply.

For these reasons, we model the unobserved heterogeneity in productivities and (possibly) preferences by allowing some of the parameters of the household welfare function to be different for households with the same observed characteristics. To do so, we adopt the latent classes modeling approach, which assumes that the population consists of a small number of different homogeneous populations or classes $K_c, c = 1, ..., C$, characterized by welfare functions with parameters $\mathbf{A}_c, \mathbf{b}_c$ (see e.g. Train (2008)). Given the probability $P(h \in K_c)$ that a household $h = 1, ..., H$ is in the class $K_c, c = 1, ..., C$, and writing now the probability that point r^* is chosen by this household, as

$$P[\Psi_{r^*} > \Psi_r, \forall r \neq r^* \mid \mu, \mathbf{A}_c, \mathbf{b}_c, \mathbf{X}] = \frac{\exp \Psi(\mu_{r^*}, \mathbf{A}_c, \mathbf{b}_c, \mathbf{X})}{\sum_{r=1}^{125} \exp \Psi(\mu_r, \mathbf{A}_c, \mathbf{b}_c, \mathbf{X})} \tag{10}$$

the unconditional probability that alternative r^* is chosen by household h is

$$\sum_{c=1}^{C} P(h \in K_c) \times P[\Psi_{r^*} > \Psi_r, \forall r \neq r^* \mid \mu, \mathbf{A}_c, \mathbf{b}_c, \mathbf{X}], \; c = 1, ..., C \tag{11}$$

Allowing for multiple latent classes makes the model more difficult to estimate, with the traditional maximum likelihood optimization methods often failing to converge. Train (2008) and Pacifico (2009) show that in such cases we can take advantage of the well-known EM algorithm. This estimation procedure is considerably faster and more stable than the traditional methods, which makes it feasible to estimate flexible models even with a large number of latent classes. This adds to the advantages of the latent class approach.

4 Data

4.1 Survey data and sample selection criteria

We estimate the models presented in the previous section on data from the Household, Income and Labor Dynamics in Australia (HILDA) Survey. The survey provides data on a wide range of socio-economic variables for a representative sample (17,000 respondents) of the Australian population, who have been followed annually since the year 2001. Particularly relevant to this study are the data on time use and the detailed information on the cost and utilization of formal and informal childcare.

Mothers with pre-school aged children represent only a small fraction of each HILDA sample. To increase sample size we construct a pooled cross-section using the four consecutive waves of HILDA from 2005 to 2008. From each wave we select partnered mothers with pre-school children. We exclude couples in which a partner is disabled, retired, or a full-time student, the husband is unemployed or the family lives in a multi-family household. We also exclude records which report incomplete or implausible survey responses.[13]

The final sample contains 1465 records. Descriptive statistics for the dependent variables and the socio-demographic characteristics entering as independent variables in X are reported in Table 1. For the purpose of comparisons by gender, the table includes descriptive statistics for male wage rates, labor supplies and time allocations to housework and home childcare.

Table 1 about here

4.2 Demographic characteristics and time use

On average, parents of pre-school children are in their early thirties, with the father around two years older than the mother. Married couples represent 83 per cent of the sample. Only 56 per cent of mothers in the sample are employed and, as we would expect, market hours distributions differ dramatically by gender, as shown graphically in Figure 1. The result is a gap of over 30 hours per week between average female and male labor supplies. The vast majority of men work full-time (more than 35 hours per week[14]) while women have a distribution of market hours that is relatively uniform, apart from a large spike at zero hours. 83 women report working over 18 hours a day for seven days a week.[15] We scale these hours to satisfy a time constraint of 18 hours per day (= 126 hours per week), retaining the relative time allocations in the original data.

Figure 1 about here

Figure 2 compares hours of housework by gender. Housework is defined to include the allocation of time to childcare as well as to errands and domestic chores. As we would expect, hours of housework are higher for females than for

[13]These mostly included households with missing data on the relevant time use variables.

[14]"Full-time employment" is defined by the Australian Bureau of Statistics (ABS) as 35 hours or more per week.

[15]The time use data are collected by questionnaire and reported as weekly time uses. Unlike diary data, questionnaire data are typically subject to larger reporting errors, and as a result the sum of individual time allocations to the various activities often fails to satisfy the time constraint.

males, as shown in Figure 3, and their leisure hours[16] are more dispersed, with substantially higher frequencies at the lower levels of weekly leisure time.

Figures 2 and 3 about here

It is clear that for this group of households with young children, the total work burden is on average larger for mothers than for fathers.

4.3 Childcare

We differentiate between "formal care" which is provided by recognized institutions, such as kindergartens and care centers, , and "informal care" provided by the husband, grandparents or other relatives, and friends. There are two reasons for this distinction. First, formal childcare differs from informal childcare in that it is recognized as incurring costs by the Australian fiscal authorities, and the family is eligible for reimbursement of a considerable part of these costs. Second, the price data on informal care is rather unreliable. The price of formal childcare is reported for all children in registered care. In contrast, informal childcare is often provided with no charge, or at a price that implies an unobserved subsidy from the carer. The lack of more detailed information about the costs of informal childcare makes any effort to impute corresponding prices infeasible. Therefore, we consider the choice of formal care only, treating informal care as exogenously given.[17] Informal care enters the utility function through X in (7), measured in hours, without a specified price.

Formal care is used by 43% of the families, while the use of informal childcare is almost universal (only 9 families report that they used no form of informal childcare). The distributions of the weekly hours of childcare are presented in Figure 4. The profiles for both types of care are relatively similar, although the formal care distribution does not go far above 60 hours per week. This reflects the fact that formal care centers are closed on weekends.

Figure 4 about here

4.4 labor income and non-labor income

Annual labor incomes are derived from reported weekly gross salaries from all jobs. The annual non-labor income of the couple is computed as the sum of each partner's reported business income, investment income, private domestic pensions and overseas pensions. Figure 5 presents distributions of male and female labor incomes and household non-labor income. According to these data,

[16] Leisure is computed as the remainder of the daily time endowment after subtracting market work and housework hours, which may be adjusted to satisfy the total time constraint. The 42-hours threshold is a consequence of this computation.

[17] An economic rationale for this would be that informal child care is quantity-rationed and has a lower cost than formal child care, the price of which determines demand for child care at the margin.

around 45% of mothers have zero labor income, while 54% of families in the sample have zero non-wage income. The distribution of non-labor income for the subsample of families with non-negative incomes is skewed towards zero. At the same time several outliers report very large incomes from business and investments.

Figure 5 about here

These income data are used to derive the set of 25 family incomes, net of the taxes and benefits and cost of childcare, associated with the discrete time use choices. All incomes are deflated to 2005, the selected base year, using the Australian consumer price index.

4.5 Family income taxes and childcare subsidies

We calculate net household income as gross income net of tax liabilities under the Personal Income Tax (PIT), Low Income Tax Offset (LITO) and Medicare Levy (ML)[18] and net of cash transfers under Family Tax Benefit Part A (FTB-A) and Family Tax Benefit Part B (FTB-B). These are the key tax policy instrument setting the parameters of the Australian family income tax system. The calculation of the net price of formal childcare takes account of the two main subsidies for childcare, Childcare Benefit (CCB) and the Childcare Rebate (CCR).

The tax base for the PIT and LITO is the individual, and therefore the marginal and average rates on the income of each partner are independent. In contrast, the tax base of the ML is partly joint income, due to the withdrawal of exemption limits on family income. Tax rates therefore become partly interdependent, with higher rates for a second earner over some range of "primary income".[19] Cash transfers under FTB-A are also withdrawn on family income. In the discussion to follow we show how these policy instruments raise tax rates on mothers as second earners by effectively replacing the individual tax base of the PIT and LITO with that of joint income, for the vast majority of families with dependent children. We also show how the LITO undermines the strict progressivity of the PIT. For the purpose of illustration, we take the rates applying in the 2007-08 financial year[20] and present results for a family with two children under 13 years, with the younger child under 5 years.

4.5.1 Personal Income Tax and LITO

The marginal rate scale of the 2007-08 PIT is strictly progressive, beginning with a zero rated threshold of $6,000, followed by rates of 15%, 30% and 40%

[18]Despite its title, the ML is entirely an income tax. It is not tied to funding any aspect of the health system.

[19]"Primary income" here refers to the income of the partner with the higher income.

[20]While the rate scale and brackets of the PIT and the LIT0 vary across the four waves of the HILDA data we are using, the structure is essentially the same in each financial year.

up to an income of $150,000, and thereafter a top rate of 45%. However, when the LITO is added, strict progressivity is lost. In 2007-08 the LITO provided a tax credit of $750, phased out at 4 cents in the dollar on individual incomes above $30,000. The net effect is a new, and less transparent, rate scale with a zero rated threshold of $11,000 and a higher rate of 34 cents in the dollar on incomes from $30,001 to $48,750, as illustrated in Figure 6. The figure plots marginal and average tax rate profiles with respect to individual taxable income. Because the marginal tax rates of partners are independent, both partners face the same marginal and average rates at any given level of income.

Figure 6 about here

The LITO was first introduced in 1993 and has been increased gradually over successive government budgets, with the effect of rasing the "middle" marginal tax rate of the PIT scale across an ever widening band of middle incomes. While expanding the LITO, the government simultaneously lowered tax rates at higher income levels over a period of significant revenue gains resulting from bracket creep (tax brackets of the PIT scale are not indexed). For example, over the period 2004-05 to 2007-08, the top marginal tax rate fell by 2 percentage points, while the income threshold for the top rate rose from $70,000 to $150,000 pa.[21] Very few working mothers gain from these tax cuts because few earn incomes at these levels.

4.5.2 Medicare Levy, FTB-A and FTB-B

The ML raises marginal tax rates by 1.5 percentage points for taxpayers with incomes above specified thresholds for exemption categories or reductions.[22] For the two-parent family the exemption threshold income is based on family income and varies with the number of children. In 2007-08 the family income limit for a full reduction for a two-parent family was $29,207, plus $2,682 for each dependent child or student. The exemption is withdrawn at a rate of 8.5 cents in the dollar above this limit, which has the effect of introducing a new marginal tax rate of 44 cents in the dollar across income bands that are well below average earnings. Thus the ML achieves a further shift away from the strictly progressive rate scale of the PIT and, at the same time, by defining the threshold for the exemption on family income, shifts the tax base towards joint income.

Family tax benefits have a more profound effect of the same kind. FTB-A provides a cash transfer for each dependent child, with the size of the transfer varying with the age of the child. The "Maximum Rate" of FTB-A in 2007-08 for a child under 13 years was $4,460.30. This maximum payment is withdrawn

[21] The combined effect of changes in the PIT scale and rising LITO over the last two and a half decades has been a significant shift in the burden of income taxation toward average earners (see Allen and Apps, 2012).

[22] There is also a surcharge for individuals and families on higher incomes who do not have private patient hospital cover – calculated at an additional 1 per cent of taxable income.

at 20 cents in the dollar on a family income over $41,318 up to the "Base Rate" of $1,890.70 pa. The Base Rate is withdrawn at 30 cents in the dollar at a higher family income threshold that depends on the number of dependent children, e.g., for a family with two dependent children, the income threshold for the Base Rate is $95,192.

FTB-B provides a payment of $3,584.30 pa for a family with a child under 5 years.[23] The payment is withdrawn at a rate of 20 cents in the dollar on a second income above $4,380. It can therefore be classified as a gender based tax[24] with, paradoxically, the higher rate applying to the income of the mother as second earner.[25]

Figure 7 compares graphically the resulting profiles of marginal tax rates with respect to primary income for a single-income family and for the second earner in a two-income family in which both partners earn the same incomes.[26] FTB payments are for two children under 13, with one under 5 years.

Figure 7 about here

The graph highlights two important effects of the Family Tax Benefit system and ML. First, they raise marginal rates across bands of second income up to the family income threshold at which the payments are fully withdrawn. Secondly, they introduce a marginal rate scale with the highest rates applying across relatively low and average incomes. For example, the marginal tax rate on a mother's earnings in a family with a primary earner on an average income can rise to over 70 cents in the dollar.

If we treat primary income as fixed and calculate the additional tax a family pays when the mother goes out to work, we obtain an average tax rate profile that includes rates of over 40 per cent, as shown in Figure 7. Consistent with joint taxation, average tax rates on the two-income family are much higher than the rates on the single-income family at any given level of primary income, until the payments are fully withdrawn. Under a tax system of this kind, the low level of female hours reported in the preceding section is hardly surprising.

4.5.3 Childcare Benefit and Childcare Rebate

Childcare Benefit depends (among other things) on the ages of children, number of children, type of childcare and the hours of childcare used. The benefit is phased out with rising family income according to the age of the child and the number of children receiving childcare.

[23] For a family with the youngest child aged 5 to 15 or full time student aged 16 to 18, the payment is $2,595.15 per year.

[24] See Alesina, Ichino and Karabarbounis (2007).

[25] Note that with selective taxation of this kind, partners in the two-income family no longer face the same marginal tax rates.

[26] Note that the marginal tax rate profile for the primary income earner in the two-income household is the same as that shown for the second earner apart from higher rates at very low levels of second income due to FTB-B.

The Childcare Rebate reimburses families for their claimed childcare expenses. It can cover up to 50% of the net childcare expenses (that is, after subtracting CCB). The CCR rate is not income-tested, but it has an upper cap on the amount of expenses which can be reimbursed. For the year 2008, this cap was $4,354 per year.

5 Results

We first present the results for the baseline homogeneous specification presented in subsection 3.1, and then discuss those for the extended model dealing with unobserved heterogeneity, as detailed in subsection 3.2.

5.1 Baseline Model: No Heterogeneity

The estimated parameters of the baseline model are reported in Table 2. While the model does not control for unobserved heterogeneity, the computation time is minimal. Moreover, if the homogeneity assumption were found to be valid, the results are consistent and more efficient than the latent class model. The coefficients indicate that several of the interaction terms yield intuitively plausible results. Increases in the number of pre-school aged children present in the household raise the marginal utility of formal childcare, and therefore strengthen the demand for it. On the other hand, increases in the (assumed exogenous) availability of informal childcare weaken it. The same is true for the time allocation to housework.

Table 2 about here

The estimated marginal utilities of the choice variables, the components of the vector μ, are central to our analysis, but their evaluation is more complex than consideration of the simple regression coefficients in isolation, since they involve the total effects of a change in one of these variables working through the entire matrix $\mathbf{A}$ and vector $\mathbf{b}$. They are also household-specific, since the utilities $\Psi(\mu)$ depend on the household's socio-demographic characteristics, $\mathbf{X}$, as well as the values of its choice variables. Therefore we present the marginal utilities in two ways: first, by averaging them across the sample of households; and secondly, by presenting the proportion of households that are measured as having negative marginal utilities for each given choice variable. It is also useful to identify separately the average marginal utility of each choice variable for the subset of households which buy formal childcare. The results are presented in Table 3.

Table 3 about here

As we expect intuitively, marginal utilities of income, domestic childcare and housework/leisure are on average positive, with only very small fractions of households having negative values at their computed optimal choice values. On the other hand, around 90% of households are reported as having negative marginal utilities of formal childcare at their optimal choice levels. This is of course not a problem for those households which choose to consume zero amounts of market childcare, but the last column of the table shows that those households consuming positive amounts also on average have negative marginal utilities, which is clearly inconsistent with it having a positive price.

This counter-intuitive result can be potentially attributed to unobserved heterogeneity. An important observation in this respect is that our sample contains a substantial share (57%) of households which do not use any formal childcare. Such sample composition can prove problematic for the homogenous model, provided that the decision about formal childcare utilization is influenced by unobserved differences in mothers' housework productivity. The model will try to explain this relation in terms of the variables included in the utility function, assigning strong disutility to formal childcare. Failure to account for unobserved heterogeneity will cause the coefficients of childcare to be biased, making it seemingly unattractive even for the active users. Therefore, it is indeed advisable to extend our analysis further and attempt to control for the unobserved heterogeneity in a flexible form.

5.2 Allowing for unobserved heterogeneity: the latent class model

A key step in the EM estimation procedure is the initial selection of the number of latent classes. This decision involves a trade-off. On the one hand, the higher the number of heterogeneous groups, the better will be the fit of the model, as we account for unobserved heterogeneity in a more flexible form. On the other hand, given that our sample is finite, more stratified models are bound to prove less efficient, as we estimate a new set of regression coefficients with each additional class. The determination of the number of classes is therefore crucial for identifying the optimal model.

Following Train (2008), we compare the models with varying classification choices on the basis of their Schwarz-Bayesian information criteria (BIC)

$$BIC = -2\log(L) + k\log(n) \tag{12}$$

where L is the likelihood, k is the number of free parameters in the model and n is the number of observations in our sample. The multiple-class models yield the following statistics:

Table 4 about here

As we see, the 8-class model attains the lowest BIC, and should therefore be considered as the most reliable specification for further analysis.

We do not present the regression coefficients for the 8-class model because the class-level stratification makes their interpretation practically infeasible. The indicators of average marginal utilities suffer from the same problem due to the logit normalization of each latent class. However, one statistic which remains readily interpretable is the share of the population with negative marginal utilities.

Table 5 about here

The only share which exhibits a substantial change compared to the baseline specification (cf. Table 3) is the one corresponding to formal childcare. The share of mothers exhibiting disutility from additional childcare drops by 30 percentage points, attaining 53% in total, and 31% when we restrict ourselves to the mothers who are actively using formal childcare.[27] This is a considerable improvement compared to the homogenous specification.

The relative performance of the models with varying numbers of latent classes is further tested through a series of simulations in the next section. The aim of these simulations is to predict how people respond to certain changes within their economic environment. By predicting (and comparing) the behavioral responses for different model specifications, we can draw inferences about the importance of class-level heterogeneity, and assess the validity of the homogeneity assumption.

6 Micro simulations

We evaluate the following changes: first, we simulate two basic adjustments to the aggregate price level - a 10% increase in the net wages of mothers, and a 10% increase in the net prices of formal childcare (Section 6.1). Second, we carry out a policy simulation in the spirit of Apps & Rees (2009), building on their critique of FTB and other joint-income fiscal measures (as discussed in the previous section). We propose an alternative system of taxes and benefits which aims to be less distortionary in respect of female labor supply than the current one, and we estimate its impact on household choices within the sample of households.

6.1 Changing net wages of mothers and net childcare prices

Increasing net wages results in higher disposable incomes for working mothers, and increasing net childcare prices has the opposite effect for families who are using formally provided childcare. The analysis of net prices and wages is advantageous over the analysis of gross indicators, as it circumvents secondary income effects caused by changes in the effective fiscal rules[28]. The resulting change in disposable incomes is hence equi-proportionate across households, which allows us to isolate the direct effect of the proportional increase of wages and childcare prices.

[27] The latter is obtained by taking weighted means over all classes, where the weights are the class probabilities given the observed choice.

[28] See Section 4.5

The impact of the pricing changes is measured in terms of aggregate elasticities. We compute the ratios of percentage changes in the relevant time use and care variables to the percentage changes in the underlying policy instruments (wages and prices) which are 10 percent, by construction. Computation of the aggregate elasticities is carried out in the following way. We first derive the pre-reform time use allocations. This is done by averaging individual choice probabilities predicted by our logit model, and multiplying the results by the hours of time use activities corresponding to the given choice. This provides us with a simulation of average time use hours, based on the estimated model.[29]

The computation of post-reform average hours is very similar to the pre-reform case. The only difference is that for the initial prediction of household choice probabilities we construct a new dataset with family incomes computed to be consistent with the specified reform. By feeding the augmented dataset into our model, we can derive choice probabilities which would correspond to the post-reform state of the world. The computation of average hours is then the same as above.

Having the pre-reform and post-reform allocations in hand, the last step in the computation of the elasticities is a matter of straightforward arithmetic. The results for models with varying numbers of classes are provided in Table 6.[30]

Table 6 about here

When we allow for latent classes, the predicted responses to the wage increases in the policy reforms fall substantially, and remain relatively stable among models with different number of classes. These results demonstrate the importance of controlling for unobserved heterogeneity in our analysis.

Standard errors are on average rising as we allow for further classes, making some of the coefficients less significant for heavily stratified models. This reflects the fact that these models are more data-intensive. Nevertheless, in certain cases we observe that the standard errors fall as we move to the more stratified models. We attribute this effect to increased explanatory power of the latter specifications.

Given the results of the BIC selection procedure discussed in the previous section, the following discussion of simulation outcomes will focus on the optimal parameterization, that is, on the 8-class model.

In the case of the net wage increase, we observe time use shifts which corre-

[29]Ideally these simulated values should be close to their empirically observed counterparts. For the 8-class model, the simulated daily averages are 1.2 hours for formal childcare, 2.5 hours for work, and 10 hours for housework. The corresponding observed mean values in the data (after replacing each observed value by the mean of its category to account for the discretization) are 1.2, 2.7 and 10.1, respectively.

[30]Standard errors on the elasticities were computed through 100 Monte Carlo simulations, re-computing the impact of the reforms with simulated sets of preference coefficients $\mathbf{A}$ and $\mathbf{b}$. The coefficients were drawn through Cholesky decomposition of an underlying covariance matrix, which was derived using a ML procedure proposed by Ruud (1991), correcting for variance of the aggregate class shares and covariance structure between different class-level parameterizations.

spond to intuition. With the reform in place, a 10% wage increase results in a 3.9% rise in average working hours, associated with a 2.7% increase in hours of bought-in childcare.

The housework elasticity proves to be negative, at -0.06. The negative sign implies that higher wages lead the mothers to work less in the household. However, the actual change of housework hours is not large enough to equilibrate the increase of working hours, which means that mothers give up both their housework and leisure activities in order to work more.

Turning to the impact of the rise in childcare prices, it is not surprising that the highest elasticity is that of formal care itself. With a 10% rise in childcare prices, the demand for formal childcare falls by 6.9%. This in turn causes the mothers to work less in the market (hours drop by 1.4%), as they have to substitute their own time for bought-in services. Accordingly, the hours of housework increase by 0.3%, replacing almost all of the forgone time formerly spent on market work.

6.2 Policy simulation: alternative system of taxes and benefits

As discussed earlier, the shift to joint taxation with the means testing of FTB-A and the ML on family income can be expected to have substantial disincentive effects on female labor supply. We analyze this by simulating the effects of switching to an individual based income tax with universal payments.[31] We construct the reform by eliminating the ML and making the payments under FTB-A universal. A key feature of the reform is the removal of the excessively high effective marginal rates on the incomes of the majority of mothers under the existing system.

To fund the increase in benefit payments we introduce a proportional increase in the marginal tax rates of the PIT & LITO (see Section 4.5.1). Assuming no behavioral responses, we calculate that an increase in tax rates of 26.76% would be required for revenue neutrality.[32] On the basis of this figure, we multiply each rate of the income tax by 1.2676, making the resulting personal income tax more progressive than the original. Figure 8 presents the differences in the net tax positions of the households in our sample induced by the reform. The differentials are ordered by the corresponding pre-reform net household incomes, so that we can see how the tax burden shifts over different income groups.[33]

Figure 8 about here

[31] As proposed in Apps and Rees (2009).

[32] The maintained assumption of no time allocation changes will inevitably be rejected by our simulation results, but it serves as a useful reference point. When we predict the reform-induced changes in work and childcare hours, we are able to calculate the gains (or losses) in tax revenue relative to the original setting due to these behavioral responses.

[33] Again, for the sake of presentation, we maintain the assumption that the individual time allocations remain unchanged.

Looking at the scatter plot, we see that the increased progressivity of the post-reform tax system shifts the tax burden towards the higher incomes, leaving middle-income families better off as compared to their original net tax positions. It is also interesting to note that the increased progressivity can be considered to be a remedy for the forgone phasing-out of the family tax benefit, because the universal FTB payments will eventually be subtracted from the incomes of high-earning households through the effect of increased income taxes.

Table 7 about here

As for the behavioral implications of our simulated policy change, Table 7 presents changes in the time allocations induced by the reform. Similarly to the previous simulations, we observe a large discrepancy between the sizes of the changes predicted by the homogenous and latent-class models, with the results of the latent-class models proving relatively stable among different specifications.

Regarding the simulation outcomes, we again restrict ourselves to the 8-class specification, in which we observe a 3.11% increase in the hours of work, a 1.75% increase of hours of formal care, and a 0.63% decrease in the hours of housework.

An interesting picture emerges when we compare responses to the FTB reform on extensive and intensive margin of the labor supply. According to our results, the observed increase of aggregate working hours can be attributed both to the newly employed mothers and to the mothers extending their original work allocation. This effect is stronger for mothers who report to work only in the household, with the absolute increase of market work hours being 28% larger than for those who are already in the labor force.

As for the change of formal childcare hours, the mothers in the household exhibit rather modest increase in absolute terms (70% lower than the employed mothers), however in relative terms they raise their childcare allocation more than employed women (the initial level of childcare utilization is substantially lower for mothers in the household).

Naturally, this behavioral heterogeneity is crucial for successful targeting of the policy reform, as it helps us to identify its actual impacts on different subsamples of the population. But it is also interesting from the perspective of economic modelling, as we can compare the predicted outcomes of homogenous and latent-class models. When we run a similar analysis of behavioral responses with the homogenous model specification, all the elasticities prove to be almost identical among both subgroups. Clearly, the homogenous model fails to predict heterogenous responses on the extensive and intensive margin of labor supply. This fallacy is further accented by the fact that the model cannot replicate observed differences in reported time use allocations among the two groups, overestimating work and formal childcare allocations of mothers in the household, and underestimating them for the subsample of employed mothers. For these reasons, it is particularly hard to maintain that homogenous model would be able to provide practitioner with unbiased predictions of the responses to proposed policy changes, as it cannot successfully replicate the behavior of heterogenous groups present in the population sample.

We further analyze the net fiscal effect of the FTB reform, accounting for subsequent changes in time allocations predicted by our 8-class model. These changes play a major role in the context of reform evaluation, because they can turn the reform either more costly or more profitable than the original fiscal system[34].

The key result in this context is that government breaks even and actually improves its net fiscal position, although the nominal change is rather marginal. The income taxes of mothers rise by mere 0.5%, compared to the 3.1% increase of aggregate working hours. Such difference in magnitudes relates to the fact that the FTB reform exerts highly heterogenous behavioral responses across households. Despite the aggregate increase of working hours among mothers, a considerable fraction (31%) actually reduce their work intensity, which lowers the expected income tax proceeds. These mothers are mainly high-earners, who are already working and whose expected income falls after introducing the reform (due to the increase of income tax rates). Because of the progression of the income tax system, the fall of revenue from high earners offsets a large part of the new tax proceeds coming from low- and middle-income households, and pulls down the aggreagate revenue effect. More specifically, the expected average increment in annual income tax proceeds is 151$ for the mothers who raise their working hours (5.6% of their initial proceeds). For the rest of the sample, the tax proceeds decline by 261$ (2.3% of their initial proceeds), lowering the aggregate impact on income taxes to 25$ on average, which translates into the aforementioned 0.5% increase of mothers' income taxes.

The situation is very similar for the childcare benefits, which also exhibit only slight increase of aggregate CCB & CCR payments as high-earning mothers opt for less formal child care[35]. The increase of annual CCB payments is rather small compared to the increase of tax proceeds, attaining on average 2$ (0.1% of the initial proceeds).

Adding the two fiscal effects together, we conclude that the households are expected to raise their annual government contribution by 23$ on average. This represents 0.2% increase of their total income tax proceeds after subtracting all the family- and childcare-related benefits. However, despite having only marginal revenue effect on the side of government, the FTB reform is successful in two other ways. It is able to increase employment among the population of mothers with preschool children, and reduce the fiscal imbalances inherent in Australian fiscal system. Furthermore, it should be emphasized that the size of

[34]Changes in time allocations can affect the government revenue through two distinct channels: by raising the work intensity the mothers are also increasing their income tax proceeds, and by choosing longer childcare hours the families become eligible for more childcare benefits.

[35]This result hinges on the fact that the CCR benefits are calculated as a proportion of the total cost of childcare, allowing users of high-priced childcare to receive larger benefits. With a system of benefits independent on claimed prices, the increase of childcare benefits would be proportional to the elasticity of formal childcare hours, that is, 1.75%.

the reform effects is heavily dependent on the degree to which is the national tax system dependent on joint taxation. Running a similar policy simulation on data from countries with purely joint-income taxation would make the behavioral responses considerably stronger.

6.3 Robustness checks

In order to assess stability of our results, we run a series of sensitivity checks, altering the econometric specification of our model in following ways:

To achieve more flexible specification, we divide the time use variables into a finer grid (6^3) of discrete choices, allowing for greater degree of discretion in the household-level decision making. We also experiment with the composition of time use variables, reducing the mother's housework decision to a single choice of motherly childcare[36]. Another pursued extension augments the model by fixed costs of working, being estimated as an additional parameter of the utility function. Furthermore, to account for potential misreporting in the individual household accounts, we estimate a model with all wages and childcare prices imputed by our model.

Table 8 about here

The elasticities presented in Table 8 confirm that changes in the econometric specification are likely to change nominal values of our results. Their relative sizes and signs remain however similar to the original model, with most of the values remaining in the 95% confidence interval of the corresponding baseline elasticities.

The stability of the elasticities is interesting in context of the model containing motherly childcare decision, as it suggests that changes in the housework allocation are proportionate, irrespective of the distinction between childcare-related and other household activities. An engagement in the labor market will hence make the women work less in the household, delegating part of their chores either to the husband, or buying in the services from the market.

Apart from direct changes of the model, we also evaluate the validity of standard errors corresponding to the measured elasticities, controlling for general heteroskedasticity and household-specific clustering. In both cases the newly derived standard errors preserve significance levels attained by the standard approach, suggesting that heteroskedasticity is not likely to distort our estimates.

[36]That way, we can examine direct substitution effects between informal and formal childcare. However, such adjustment comes at the cost of making the residual time allocation more difficult to interpret, as it contains both leisure time and housework unrelated to childcare

7 Conclusions

In this paper we have analyzed the time allocation decisions of mothers with pre-school children with emphasis on their labor supply choices. We have focused on the identification and analysis of unobserved heterogeneity, which has its origins in across-household variations in productivity and preferences and which has proven to play a dominant role in the decision making of mothers in our data. The heterogeneity in unobserved productivities and tastes among the identified latent groups undermines the usefulness of the homogenous model with no unobserved heterogeneity. The parameters fail to capture the true effects of factors driving household decision making, and hence simulations based on the baseline homogeneous model specification can be expected to give misleading results.

To control for unobserved heterogeneity, we estimated a series of latent-class multinomial logit models, taking the 8-class model to be the best parameterization. This model was found to perform optimally, balancing goodness of fit on the one hand against parsimony on the other. To assess the responsiveness of our sample to changes in the tax system and in childcare prices, we conducted a series of policy reform simulations, increasing net wages of mothers and net childcare prices in the first two reforms, and altering the joint-income structure of the existing system in the third reform.

The first two micro simulations based upon our estimates show that mothers are responsive both to the changes in wages and childcare prices. This result suggests that market work and formal childcare tend to be complements, and respond significantly to wage and price changes. The results also indicate that significant changes in labor supply and childcare demand can remain unidentified when the unobserved heterogeneity is left untreated, thus significantly distorting the size of predicted changes in female labor supply.

In the third simulation, we show that the tax measures that withdraw benefits on the basis of joint income are likely to prove adverse to the labor supply decisions of working mothers, and that the tax system can be made more favorable for employed mothers by switching to a fully individual based system. In such a setting, women are predicted to increase their labor supply and to utilize more formal childcare. These responses also raise additional tax revenue which could be used to lower tax rates and therefore achieve efficiency gains.

A number of improvements and extensions are of course possible. First, our analysis would benefit from exploiting the panel structure of the HILDA dataset, controlling for time-stable individual effects. Secondly, although we consider the current method of treating unobserved heterogeneity to perform well, it could be worthwhile to assess the stability of our results by using alternative ways of controlling for unobserved heterogeneity, such as the random coefficient mixed logit model, or the approaches utilizing Bayesian nonparametric methods.

References

[1] ALESINA, A., A. ICHINO & L. KARABARBOUNIS (2007): "Gender based taxation and the division of family chores". Harvard University Discussion Paper.

[2] APPS, P. & D. ALLEN (2012): TBA, forthcoming

[3] APPS, P. & R. REES (2009): *Public Economics and the Household*. Cambridge University Press

[4] BAKER, M., J. GRUBER & K. MILLIGAN (2008): Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy* 116(4): pp. 709–745

[5] BERNDT, E., B. HALL & R. HALL (1974): "Estimation and inference in nonlinear structural models". In *Annals of Economic and Social Measurement*, Volume 3, number 4," NBER, pp. 103-116. National Bureau of Economic Research, Inc

[6] BHAT, C. (2000): "Flexible model structures for discrete choice analysis." In *Handbook of Transport Modeling*, volume 1, pp. 71-90. Pergamon, Amsterdam, NL

[7] BLAU, D. (2003): Child care subsidy programs. In Means-Tested Transfer Programs in the United States, NBER Chapters, pp. 443–516. National Bureau of Economic Research, Inc

[8] BLUNDELL, R. & A. SHEPHARD (2011): Employment, hours of work and the optimal taxation of low income families. *IZA Discussion Papers 5745*, Institute for the Study of Labor (IZA)

[9] BURDA, M., M. HARDING & J. HAUSMAN (2008): "A Bayesian mixed logit & probit model for multinomial choice." *Journal of Econometrics* 147(2): pp. 232-246

[10] CAMERON, A. C. & P. K. TRIVEDI (2005): *Microeconometrics.* Cambridge University Press

[11] CONNELLY, R. (1992): "The Effect of Child Care Costs on Married Women's Labor Force Participation." *The Review of Economics and Statistics* 74(1): pp. 83-90

[12] CONNELLY, R. & J. KIMMEL (2003): Marital status and full-time/part-time work status in child care choices. *Applied Economics* 35(7): pp. 761–777

[13] DEMPSTER, A., N. LAIRD & D. RUBIN (1977): "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society. Series B* 39(1): pp. 1-38

[14] DOIRON, D. & G. KALB (2005): Demands for child care and household labour supply in Australia. *The Economic Record* 81(254): pp. 215–236

[15] DUNCAN, A., G. PAULL & J. TAYLOR (2001): Mothers' employment and the use of childcare in the uk. *IFS Working Papers W01/23*, Institute for Fiscal Studies

[16] HECKMAN, J. (1979): "Sample selection bias as a specification error." *Econometrica* 47(1): pp. 153-161

[17] HECKMAN, J. & B. SINGER (1984): "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica* 52(2): pp. 271-320

[18] KALENKOSKI, C. M., D. C. RIBAR & L. S. STRATTON (2005): Parental child care in single-parent, cohabiting, and married-couple families: Time-diary evidence from the united kingdom. *American Economic Review* 95(2): pp. 194–198

[19] KEANE, M. & R. MOFFITT (1998): "A structural model of multiple welfare program participation and labor supply." *International Economic Review* 39(3): pp. 553-589

[20] KIMMEL, J. & R. CONNELLY (2007): Mothers' time choices: Caregiving, leisure, home production, and paid work. *The Journal of Human Resources* 42(3): pp. pp. 643–681

[21] KORNSTAD, T. & T. THORENSEN (2007): A discrete choice model for labor supply and childcare. *Journal of Population Economics* 20(4): pp. 781–803

[22] PACIFICO, D. (2009): "On the role of unobserved preference heterogeneity in discrete choice models of labor supply." *Working Paper Univ. Modena* 1

[23] RIBAR, D. C. (1995): A structural model of child care and the labor supply of married women. *Journal of Labor Economics* 13(3): pp. 558–97

[24] RUUD, P. (1991): "Extensions of estimation methods using the EM algorithm." *Journal of Econometrics* 49(3): pp. 305-341

[25] TRAIN, K. (2008): "EM algorithms for nonparametric estimation of mixing distributions." *Journal of Choice Modelling* 1(1): pp. 40-69

[26] TRAIN, K. (2009): *Discrete Choice Methods with Simulation.* Cambridge University Press

[27] VAN SOEST, A. (1995): "Structural models of family labor supply: A discrete choice approach." *Journal of Human Resources* 30(1): pp. 63-88

[28] VAN SOEST, A. & E. STANCANELLI (2010): "Does Income Taxation Affect Partners Household Chores ?" *IZA Working Paper* 1(5038)

[29] VAN SOEST, A., A., M. DAS & X. GONG (2002): A structural labour supply model with flexible preferences. *Journal of Econometrics* 107(1-2): pp. 345–374

Figures and Tables

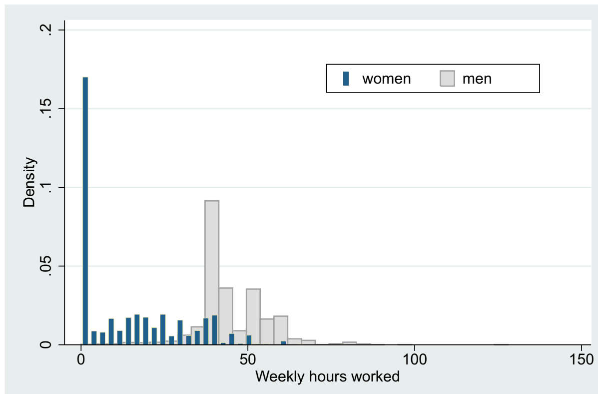

Figure 1: Distribution of weekly hours of market work in families with preschool children

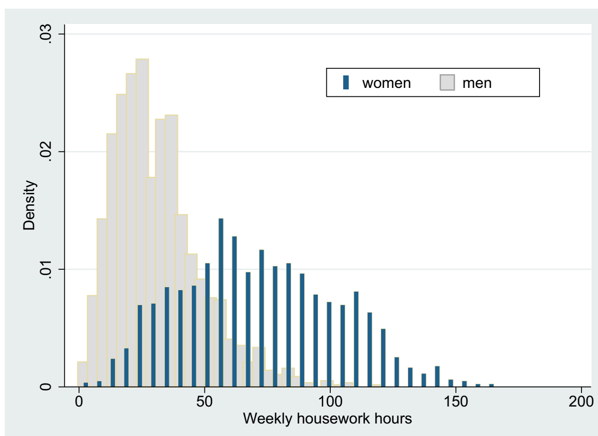

Figure 2: Distribution of weekly hours spent on housework in families with preschool children

.

.

.

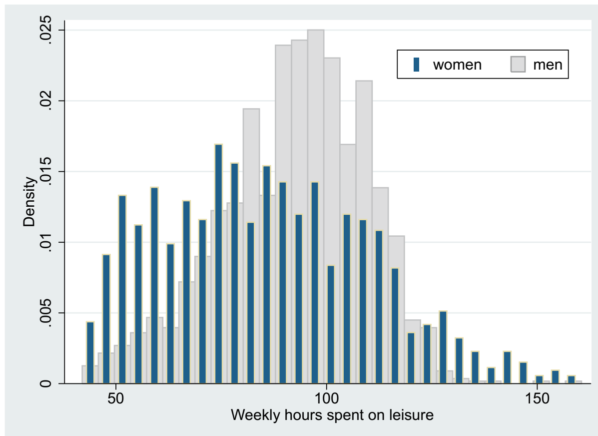

Figure 3: Distribution of weekly hours spent on leisure in families with preschool children

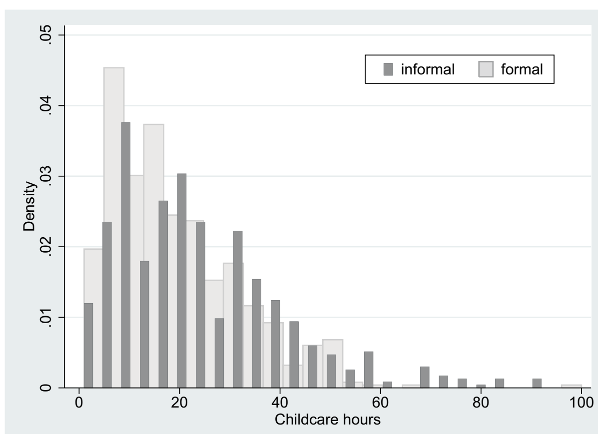

Figure 4: Distribution of weekly hours of informal and formal childcare, families with preschool children who are using childcare

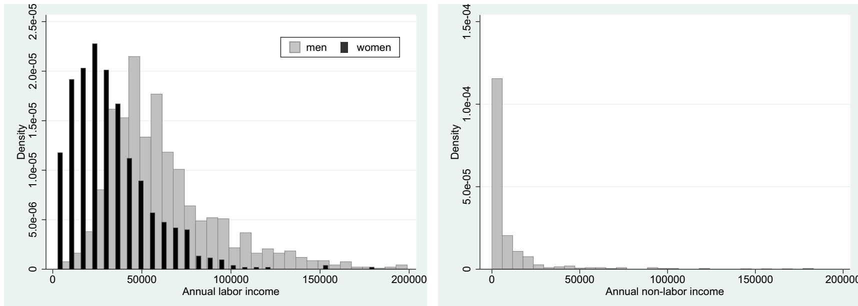

Figure 5: Distributions of labor and non-labor gross annual income, in Australian dollars, families with preschool children

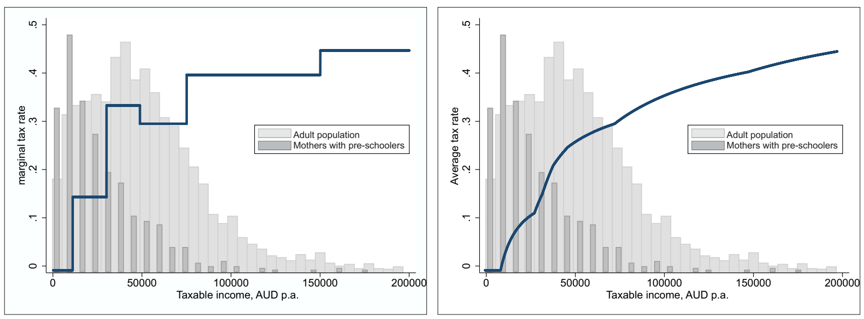

Figure 6: Marginal and average income tax profiles, and annual income distributions for year 2008

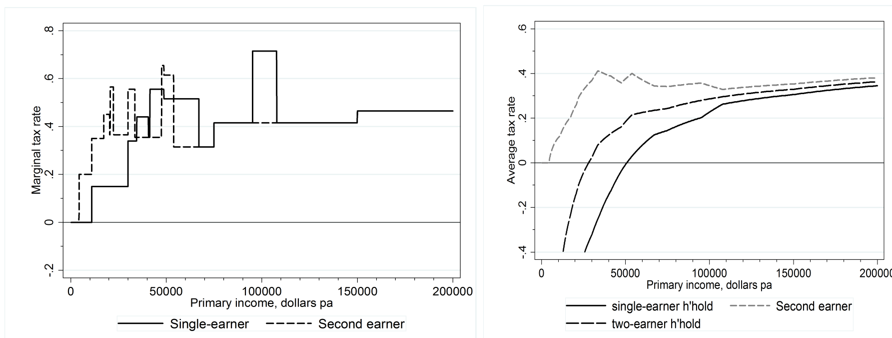

Figure 7: Marginal and average income tax profiles for one-earner and two-earner households, including income tax measures and family tax benefits

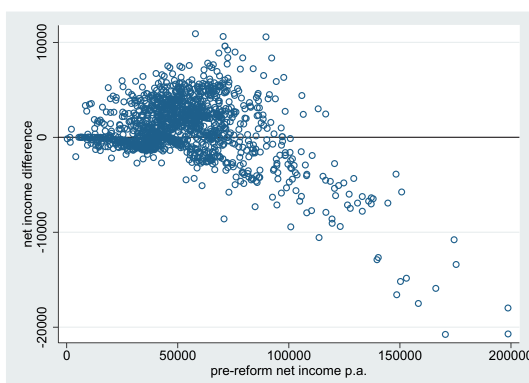

Figure 8: Post-reform differences in the net tax positions of the families, ordered by their pre-reform net incomes.

Table 1: Summary statistics, sample of couples with preschool children

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Age - women	32.9	5.6	16	48
Age - men	34.9	6.3	18	58
Marital status (dummy)	0.82	0.4	0	1
Employment status - women (dummy)	0.55	0.5	0	1
Employment status - men (dummy)	1	0	1	1
Number of children, aged 0-4	1.39	0.6	1	4
Number of children, aged 5-9	0.47	0.7	0	4
Number of children, total	2.01	0.96	1	6
Market work - women, weekly hours	13.9	15.8	0	80
Market work - men, weekly hours	44.5	11.3	0	128
Informal childcare - women, weekly hours	41.2	25.8	0	128
Informal childcare - men, weekly hours	16.4	11.6	0	80
Informal childcare - other relatives & friends, weekly hours	6.9	13.6	0	120
Housework (excl. childcare) - women, weekly hours	30.4	18.0	0	133.3
Housework (excl. childcare) - men, weekly hours	14.7	10.9	0	80
Leisure - women, weekly hours	82.3	27.6	0	160
Leisure - men, weekly hours	91.9	18.01	10	150.5
Formal childcare, weekly hours	8.3	12.8	0	100
Formal childcare price, in AUD, hourly	8.7	3.5	1.583	23.8
Annual wage - men, in AUD	63373	35736	5136	357216
Annual wage - women, in AUD	18514	23570	0	182256
Annual non-labour family income, in AUD	6617	31778	0	683974
Number of observations		1465		

Table 2: Regression results for the baseline homogenous model

Matrix A		Vector b	
income	.199 (.092)**	formal care*log(age)	-.613 (3.63)
formal care	2.23 (6.20)	formal care*log(age)^2	.130 (.529)
housework	4.71 (3.42)	formal care*married	-.031 (.057)
leisure	7.59 (3.52)**	formal care*No. dependent children	-.127 (.037)***
formal care^2	.128 (.014)***	formal care*children aged 0-4	.329 (.056)***
housework^2	.066 (.005)***	formal care*children aged 5-9	.093 (.049)*
leisure^2	.022 (.006)***	formal care*informal care	-.009 (.001)***
inc*formal care	.011 (.004)**	housework*log(age)	-4.07 (2.01)**
inc*housework	.019 (.002)***	housework*log(age)^2	.607 (.294)**
income*leisure	.020 (.002)***	housework*married	.026 (.032)
formal care*housework	-.066 (.004)***	housework*No. dependent children	-.112 (.020)***
formal care*leisure	-.064 (.005)***	housework*children aged 0-4	.336 (.032)***
housework*leisure	-.050 (.005)***	housework*children aged 5-9	.148 (.027)***
		housework*informal care	-.008 (.001)***
		leisure*log(age)	-4.87 (2.07)**
		leisure*log(age)^2	.685 (.305)**
		leisure*married	.012 (.034)
		leisure*No. dependent children	-.053 (.023)**
		leisure*children aged 0-4	.231 (.034)***
		leisure*children aged 5-9	.114 (.031)***
		leisure*informal care	-.009 (.001)***
n			1465
log-likelihood			-6076.44

standard errors in the brackets, significance levels: 90*, 95**, 99***

	average marginal utility		**negative fraction**	
	full sample	childcare users	full sample	childcare users
Income	1.06	1.04	0	0
Formal care	-0.69	-0.05	0.83	0.58
Housework	0.63	0.26	0.17	0.37
Leisure	0.64	0.33	0.24	0.28

Table 3: Average marginal utility of the main regressors and fraction of the population sample with negative marginal utilities, homogenous model

No. of classes	Log-likelihood	BIC
1	-6076.44	12415.31
2	-4921.44	10367.73
3	-4676.13	10139.53
4	-4226.96	9503.61
5	-3963.88	9239.88
6	-3748.64	9071.83
7	-3584.41	9005.79
8	**-3319.71**	**8738.83**
9	-3236.49	8823.87

Table 4: Bayesian Information Criteria for multi-class models

	whole sample	childcare users
Income	0	0
Formal care	0.53	0.31
Housework	0.24	0.41
Leisure	0.23	0.38

Table 5: Fraction of the population sample with negative marginal utilities of the main regressors, 8-class model

Mothers' net wage increased by 10%

No. of classes	1	2	3	4	5	6	7	8	9
Formal care hrs.	1.01	0.58	0.58	0.58	0.59	0.50	0.51	0.42	0.49
	(0.051)**	(0.042)***	(0.096)***	(0.101) ***	(0.070)***	(0.078)***	(0.095)***	(0.065)***	(0.097)***
Work hrs.	1.35	0.77	0.65	0.48	0.65	0.62	0.52	0.43	0.68
	(0.043)***	(0.034)***	(0.084)***	(0.088)***	(0.067)***	(0.126)***	(0.134)***	(0.086)**	(0.149)***
Housework hrs.	-0.23	-0.10	-.10	-0.06	-0.10	-0.07	-0.10	-0.08	-0.09
	(0.012)***	(0.011)***	(0.020)***	(0.016)***	(0.018)***	(0.015)***	(0.023)***	(0.015)***	(0.023)***

Net childcare price increased by 10%

No. of classes	1	2	3	4	5	6	7	8	9
Formal care hrs.	-0.51	-0.51	-0.48	-0.52	-0.47	-0.45	-0.50	-0.42	-0.38
	(0.017)***	(0.02)***	(0.031)***	(0.043)***	(0.033)***	(0.079)***	(0.06)***	(0.081)***	(0.06)***
Work hrs.	-0.17	-0.11	-0.15	-0.10	-0.12	-0.10	-0.11	-0.08	-0.11
	(0.01)***	(0.011)***	(0.018)***	(0.026)***	(0.02)***	(0.023)***	(0.027)**	(0.021)***	(0.027)***
Housework hrs.	0.03	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.01
	(0.003)***	(0.004)***	(0.006)***	(0.005)**	(0.007)***	(0.008)***	(0.008)***	(0.008)***	(0.008)***

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Elasticities of the time use allocations with respect to the changes in net wages and net childcare prices

No. of classes	1	2	3	4	5	6	7	8	9
Formal care hrs.	4.67%	2.77%	2.77%	3.52%	3.78%	1.57%	2.24%	1.75%	2.74%
	(0.243)***	(0.234)***	(0.648)***	(0.634)***	(0.54)***	(0.422)***	(0.511)**	(0.38)***	(0.564)***
Work hrs.	7.70%	4.43%	3.54%	4.16%	3.96%	2.99%	2.77%	3.11%	3.41%
	(0.337)***	(0.262)***	(0.627)***	(0.616) ***	(0.62)***	(0.83)***	(0.885)**	(0.684)***	(0.915)***
Housework hrs.	-1.31%	-0.6%	-0.57%	-0.63%	-0.39%	-0.38%	-0.66%	-0.63%	-0.61%
	(0.091)***	(0.076)***	(0.155)***	(0.113)***	(0.144)***	(0.136)***	(0.191)**	(0.133)***	(0.152)***

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Percentual changes in time allocations after the FTB reform

Mothers' net wage increased by 10%					
	original model	6 brackets	motherly care	fixed costs	imputed
Formal care hrs.	0.42 (0.065)***	0.51	0.43	0.56	0.36
Work hrs.	0.43 (0.086)**	0.61	0.57	0.64	0.47
Housework hrs.	-0.08 (0.015)***	-0.12	-0.07	-0.06	-0.06

Net childcare prices increased by 10%					
	original model	6 brackets	motherly care	fixed costs	imputed
Formal care hrs.	-0.42 (0.081)***	-0.45	-0.44	-0.41	-0.40
Work hrs.	-0.18 (0.021)***	-0.11	-0.09	-0.12	-0.10
Housework hrs.	0.02 (0.008)***	0.02	0.01	0.02	0.01

FTB reform - percentual changes					
	original model	6 brackets	motherly care	fixed costs	imputed
Formal care hrs.	1.75% (0.380)***	1.94%	2.29%	3.17%	1.92%
Work hrs.	3.11% (0.684)***	3.18%	3.01%	4.58%	3.25%
Housework hrs.	-0.61% (0.133)***	-0.77%	-0.37%	-0.51%	-0.58%

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Robustness check - elasticities and reform responses derived by alternative model specifications with 8 latent classes