Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Jan Kabátek

A Structural Model of Female Labour Supply in the Context of Childcare and Unobserved Heterogeneity

Research Master Thesis 2011-028

Tilburg University

Tilburg School of Economics and Management

Department of Econometrics & OR

MASTER THESIS

A structural model of female labour supply in the context of childcare and unobserved heterogeneity

Author: **Jan Kabátek**

Supervisor: **Prof. A.H.O. van Soest**

Academic Year: **2010/2011**

Declaration of Authorship

Hereby I declare that I have written this thesis independently, using only the listed resources and literature.

Tilburg, September 14, 2011

Signature

Acknowledgments

First and foremost, I would like to express my highest gratitude to my supervisor, professor Arthur van Soest, for allowing me to participate on a research project which has eventually turned into this thesis. I have greatly benefited from his guidance throughout the both years of my assistantship, and I am glad that our cooperation will continue even throughout my doctoral studies.

I am also grateful to the initiators of our research project, professors Patricia Apps, and Ray Rees, who have provided me with many insights from the economic theory, and who have shown immense hospitality during my stay in Sydney.

Furthermore, I would like to thank all the colleagues with whom I have consulted my work, starting with my ever-supportive friend Luc Bissonette, and continuing with David Allen, David Vonka, Hanka Vonkova, Roger Willkins, and many of my Research Master classmates at the CentER Graduate School.

Last but not least, I would like to thank my dear family for their unfailing love and care.

Abstract

This thesis aims to provide a structural model of female labor supply, tailored to accommodate labor decisions of married and cohabiting mothers with children in the preschool age. The presented model follows tradition of discrete choice time-use models, with the individual agents deciding on allocation of their available time within a set of alternative activities. Our analysis benefits from several novel estimation approaches, with the main innovation being an inclusion of formal childcare into the time-use choice set. This is made possible through the HILDA dataset, a rich micro-level survey of Australian population, which contains detailed information on individual childcare utilization and corresponding prices. Our model also explicitly controls for unobserved heterogeneity, using EM algorithm to identify latent classes of agents within our population sample. The resulting policy simulations show that the women with preschool children are highly sensitive to changes in wages and costs of childcare, adjusting their optimal time-use allocations to account for the given policy reform. Further simulations prove that the labor conditions of employed mothers are likely to be improved in a fiscal system which is based solely on the individual income taxation, in comparison to the current Australian system which combines both individual and joint-income fiscal measures.

JEL Classification	F36, F55, H20, H21, H26
Keywords	Female Labour supply, Structural modeling, EM algorithm

Author's e-mail	j.kabatek@uvt.nl
Supervisor's e-mail	avas@uvt.nl

Contents

List of Tables **vii**

List of Figures **viii**

1 Introduction **1**

2 Baseline model specification **4**

2.1 Specification of the main variables 4

2.1.1 Time-use variables 4

2.1.2 Income indicator 5

2.2 Auxiliary Heckman models 5

2.3 Utility function 6

2.4 Multinomial logit model 7

3 Unobserved heterogeneity in the labor supply models **9**

3.1 Parametric random coefficients model 10

3.2 Alternative approach - latent class models 11

3.3 EM algorithm 12

4 Data **16**

4.1 Description of the dataset and sample selection criteria 16

4.2 Description of the variables 17

4.2.1 Basic characteristics and time-use indicators 18

4.2.2 Childcare variables 20

4.2.3 labor income, non-labor income 21

4.3 Australian personal tax & social benefit system 23

4.3.1 Social benefits 24

5 Results **27**

5.1 Baseline model, no heterogeneity 27

5.2 Allowing for unobserved heterogeneity through the EM-algorithm 30
5.2.1 Model with two latent classes 31
5.2.2 Identification of the optimal number of classes 33
5.3 Policy simulations 34
5.3.1 Changing the gross wages and gross childcare prices . . . 34
5.3.2 Alternative system of taxes and benefits 37

6 Conclusion **40**

Bibliography **43**

List of Tables

4.1 Summary statistics, sample of couples with preschool children . 17

5.1 Regression results-baseline model, and model with 2 lat. classes 28
5.2 Average marginal utility of the main regressors and fraction of the population with negative marginal utilities, homogenous model 29
5.3 Marginal utility of the main regressors and fraction of the population with negative marginal utilities, 2-class model 31
5.4 Bayesian Information Criteria for multi-class models 33
5.5 Elasticities of the time-use allocations with respect to the changes in gross wages and gross childcare prices 35
5.6 Percentual changes in time allocations after the FTB reform . . 38

List of Figures

4.1 Distribution of weekly hours worked in families with preschool children 18
4.2 Distribution of weekly hours spent on housework in families with preschool children 19
4.3 Distribution of weekly hours spent on leisure in families with preschool children 19
4.4 Distribution of weekly hours of informal and formal childcare, families with preschool children who are using childcare 21
4.5 Distributions of labor and non-labor gross annual income, in Australian dollars, families with preschool children 22
4.6 Individual income tax profiles, and annual income distributions 24
4.7 Individual income tax profiles for one-earner and two-earner households, including income tax measures and family tax benefits 25

5.1 Post-reform differences in the net tax positions of the families, ordered by their pre-reform net incomes. 38

Chapter 1

Introduction

An estimation of labor supply models has been on the frontier of empirical economic research for several decades, standing out for its prominent policy importance, and extensive data availability. The efforts to approximate the complex behavioral systems underlying individual employment decisions have established a dynamic research field, which has been successfully evolving ever since[1].

And as all the major fields of economic research, also the labor supply modeling has been subject to several shifts of the estimation paradigm, with the last one involving an abandonment of sophisticated top-down models, and calling for greater flexibility and simplification of the estimation process. These appeals gained substantial momentum within the domain of structural modeling, which can be regarded as a hallmark case of flexible econometric estimation.

The structural models are characteristic for their bottom-up approach, basing macroeconomic predictions on micro-level data. That way, the econometrician does not need to impose excessive assumptions on the economic aggregates, which had been one of the main critiques of the older labor supply models. These favorable properties have led to a surge of applications in the labor supply literature, with the prominent examples being Van Soest (1995), Keane & Moffitt (1998), or Blundell & Shephard (2008).

Apart from the emphasis on flexibility and simplification, another pronounced aspect of the current labor supply models has been an individual heterogeneity. The literature shows[2] that the individual heterogeneity (be it observed or unobserved) is likely to play a substantial role in the labor supply

[1] For a thorough overview of labor supply literature, see Blundell & MaCurdy (1999)
[2] Ibid

decisions, and as such, it should be accounted for by the models.

As for the observed heterogeneity, recent work of Apps & Rees (2009) presents a sweeping critique of labor supply models treating people as independent homogenous agents. The authors claim that it is crucial to assess individual labor decisions in the context of household, accounting for characteristics of spouses, presence of children, and availability of childcare.

However, even with the properly identified household composition, the structural labor supply models are likely to face problems of dissimilar latent preferences among individuals, commonly known as the unobserved heterogeneity. This long-recognized issue has led the econometricians to augment the structural models with different controlling mechanisms, with the most important approaches being random coefficient models (used in Van Soest, 1995 or Keane & Moffitt, 1998), and latent class models (used in Burda *et al.*, 2008 or Train, 2008).

In this thesis, we are trying to combine the above mentioned developments of the labor supply literature, and estimate a structural discrete choice model in the spirit of Van Soest (1995) and Van Soest & Stancanelli (2010), focusing on labor decisions of married and cohabiting mothers with children in the preschool age.

Our labor supply model is based on a concept of time-use, with the individual agents deciding how to allocate their time during the day. Choosing from a discrete set of alternative time allocations, they divide their day into work, housework, and leisure activities. That way, the work indicator can accommodate different intensity levels, so that the model is capable of identifying more subtle changes in the labor supply, compared to the models based on binary employment decisions.

We are using a rich micro-level dataset which enables us to identify heterogeneous household characteristics, as emphasized by Apps & Rees (2009), including detailed information on utilization of childcare, and formal childcare prices. Based on the results of our model, we aim to provide consistent predictions of labor supply shifts in the context of fiscal reforms and price level changes.

The rest of this thesis is structured as follows; Chapter 2 introduces the theory underlying our baseline econometric model. Chapter 3 elaborates on the concept of unobserved heterogeneity and extends our model to control for

this data issue. Chapter 4 describes the dataset used for our estimations, and outlines the system of legislative measures which is utilized in our analysis. Chapter 5 presents the results of our estimations, focusing mainly on policy simulations based on the optimized models. Chapter 6 concludes.

Chapter 2

Baseline model specification

In this section we are going to describe building blocks of our econometric model, focusing on the baseline case without unobserved heterogeneity. We start with specification of the main variables used in our analysis, and we present how they enter individual utility contribution. Then, we describe auxiliary models which are used to derive price indicators that complement some of these variables. Lastly, we present general multinomial logit model, which constitutes the baseline estimation procedure of our analysis, and which we further extend to account for the unobserved heterogeneity.

2.1 Specification of the main variables

2.1.1 Time-use variables

We build our model around three main variables which represent alternative time allocations of the mothers. These are, hours of work, hours of housework, and hours of paid childcare.

Such duration variables are often recorded in continuous time format, so that to utilize discrete choice methods, researchers have to approximate the timeline by a finite set of intervals. In the context of this analysis, we can fragmentize each of our three duration variables into k brackets, which yields a $k \times k \times k$ grid of compound time allocations[1].

This grid reflects how can an agent utilize her time within the day (and how many hours of childcare she can buy in), which means that each grid-cell

[1] We choose k to be 4, which leaves us with a 64-cell grid of time allocations. Determination of threshold values for each of the brackets is based on corresponding sample quartiles, as observed in the original continuous-time data.

represents a unique way how to allocate one's time, choosing one of the k intensity levels for all the three activities. Now, to account for the comprehensive decision about daily time allocation, we assume that the rest of the 24 hours after subtracting paid work and housework is used on leisure[2].

2.1.2 Income indicator

Derivation of the other key variable, net household income, is tightly connected with the time-use indicators introduced in the previous section. These time allocation choices directly impact financial situation of the family, as their income is dependent both on the intensity of mother's paid work, and also on the amount of bought-in childcare. Such dependency constitutes an alternative-specific value of the net household income for each of the 64 alternative time allocations. A closer inspection of the derivation process follows:

The net household income indicator is based on the gross household income, which consists of salary of the wife $w_w t_w^w$ (derived as a product of choice-specific hours of work and her predicted wage), and the remaining household income y_h, that is, non-wage income of the couple and salary of the husband (both assumed exogenous).

The gross income is then adjusted for family-specific income taxes and social transfers T, and formal child care expenses $fc \cdot p_{fc}$ (derived as a product of choice-specific hours of child care and its predicted price), which leaves us with the net household income indicator y,

$$y = f(y_h, w_w, t_w^w, fc, p_{fc}) = y_h + w_w t_w^w - T(y_h, w_w, t_w^w, fc, p_{fc}) - fc \cdot p_{fc}. \quad (2.1)$$

2.2 Auxiliary Heckman models

In the income equation (2.1), we are utilizing two price indicators, that is, market wage of mothers, and price of formal child care. Both indicators are inherently individual specific, as the mothers differ in their productivity, socio-economic factors, or even a geographical location. Therefore, proper identification of the individual prices is crucial for the validity of our analysis.

However, since not every mother is using formal childcare services, and even

[2]The leisure is considered here in a rather broad sense, as it contains all the activities unrelated to work, including the time spent sleeping.

more importantly, since many mothers decide not to work, substantial fraction of the price information remains unobserved. These values therefore have to be imputed, drawing on the price information of the women who are actively engaged in such activities. In this respect, we follow related labor supply literature (Connelly, 1992), and employ Heckman selection models (Heckman, 1979) to model missing information in the both price indicators.

The Heckman selection model is an estimation procedure which consists of two interdependent stages, with the first one being a binary probit model of participation choice (*e.g.,* decision to work), and the second one analyzing a related continuous variable (*e.g.,* market wage) in the context of a set of regressors. The main advantage of this approach is that it corrects for potential selection bias of the second stage estimator by linking it to the first stage participation choice[3].

To avoid problems with multicollinearity, it is preferred to have the first stage of the Heckman selection model estimated with a genuine selection variable, which affects the participation decision, but not the censored indicator itself (Cameron & Trivedi, 2005). In the case of wage modeling we choose the selection variables to be the non-wage income of the family and number of children in the household. For the childcare price estimation, we choose distance to grandparents and number of other adults in the household, which follows the rationale of Connelly (1992).

2.3 Utility function

Following Van Soest & Stancanelli (2010) we model the underlying individual utility to take on the following quadratic functional form[4],

$$V(\mu) = \mu' A \mu + b' \mu \quad , \mu = (t_w^{hw}, t_w^l, fc, y), \tag{2.2}$$

where the vector of variables μ contains respectively hours of housework, hours of leisure, hours of formal care, and income.

It is important to note here that the hours of paid work are not entering

[3]This is done by including a first stage correction term called inverse Mills ratio among the set of second stage regressors.

[4]We have also experimented with other functional forms, such as CES, or Box-Cox utility specification. These however did not prove to perform better than the provided quadratic model.

this vector of variables directly. We have already established that the leisure is considered a complementary activity to the paid work and the housework, so that together the three are always forming the daily endowment of 24 hours. As a result, these variables are linearly dependent, and direct inclusion of all of them would cause perfect collinearity. Hence one of the three activities has to be left out to ensure identification.

Regarding the regression parameters entering the utility function, A is a symmetric 4x4 matrix containing quadratic coefficients of the time-use variables and income, and b' is a vector of their linear coefficients and interactions. The time-use variables in linear form are interacted with a set of individual socio-economic variables, so that

$$b_i = \sum_k \beta_k X_{ki} \tag{2.3}$$

where X_i includes a constant term, age, age squared, number of preschool-aged children, number of school-aged children, and hours of informal child care (provided by relatives, friends or the husband).

2.4 Multinomial logit model

A Multinomial logit model is the main building block of our estimation. We consider mother's choice between the 64 time allocation alternatives, and to enable empirical identification of the corresponding choice probabilities, we add i.i.d. extreme value errors ε_j into the utility evaluation of each alternative,

$$V_j(\mu_j) = V(t^{hw}_{wj}, t^{l}_{wj}, fc_j, y_j) + \varepsilon_j \quad j = 1...64 \tag{2.4}$$

Now, we can write the individual conditional probability of choosing preferred time allocation j in the usual way,

$$P_i(J = j) \;=\; P(V_j > V_k \,\forall k \neq j \,|\mu, A, b_i) = \frac{\exp\left(V(\mu_j, A, b_i)\right)}{\sum\limits_{k=1}^{64} \exp\left(V(\mu_k, A, b_i)\right)}, \tag{2.5}$$

so that the logarithmic form of this individual probability constitutes a log-likelihood contribution which is then summed over the sample of mothers and

maximized through the Maximum Likelihood estimation procedure,

$$\mathcal{L} = \sum_{i=1}^{n} \log \prod_{j=1}^{J} P(J = j|\mu, A, b_i)^{d_{ij}} \tag{2.6}$$

$$\theta^* = \arg\max_{\theta} \mathcal{L} \ , \tag{2.7}$$

where d_{ij} equals 1 if the agent i chooses the option j, and 0 otherwise.

As a numerical routine for optimizing the likelihood we have selected BHHH algorithm (Berndt *et al.*, 1974), with the gradient being computed using numerical derivatives of the log-likelihood function.

Chapter 3

Unobserved heterogeneity in the labor supply models

Unobserved heterogeneity is a widely-recognized attribute of labor supply models[1], which is posing an ongoing challenge for the practicing econometricians. One of the main concerns in this respect is a potential bias of the regression estimates derived from data stricken by the unobserved heterogeneity.

To illustrate such issue, we should point out that different groups of agents within the selected population sample are likely to have specific attributes which are not observed by the econometrician. Now, if their observed choices are (partly) driven by these unobservables, then the model with homogenous preferences can fail to capture the true effects of the regressors. This bias then spreads further, distorting any predictions or simulations based on the homogenous model, and hence can lead to adverse policy recommendations.

Another potential problem related to the unobserved heterogeneity is an assumption of independence of irrelevant alternatives (IIA), which is a characteristic attribute of the homogenous multinomial logit models (Bhat, 2000). This property makes practicioner assume that irrespective of changes in the choice set, agents will select the original alternatives in a way that retains their relative population shares.

The major fallacy of such assumption is that it disregards any differences in substitution patterns among specific choices. These patterns are often highly heterogenous, so that the relative shares of alternatives are likely to change based on the unobserved tastes of agents within the population sample. There-

[1]For an extensive survey, see Blundell & MaCurdy (1999).

fore, in order to model correct response to potential changes in the choice set, it is essential to allow for the effects of unobserved heterogeneity.

3.1 Parametric random coefficients model

Several alternative approaches were developed to overcome the unfavorable properties of homogenous preferences, with the most prominent method being parametric random coefficients model. In this model, the regression parameters are allowed to have a random, individual-specific part which is distributed according to an underlying distribution, and correlated within agents. The random part of the parameter hence accommodates for systematic deviations from the average preference, and the corresponding covariance matrix takes into account incidence of these deviations in the context of remaining random coefficients. That way, the coinciding preference outliers get bundled together, which results in better identification of the underlying heterogenous types, and facilitates modeling of the resulting behavior. Previous applications of the random coefficient approach in the labor supply literature include Van Soest (1995) or Keane & Moffitt (1998).

However, despite the long-lasting popularity of the parametric random coefficients model, its use is nowadays becoming criticized for often non-justifiable assumptions imposed on the distribution of random coefficients (see *e.g.* Burda *et al.*, 2008, or Train, 2008). These 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 preference ordering.

To illustrate the potential fallacies of unimodality assumption, we will focus on the multivariate normal case. The multivariate normal distribution is centered around the average value of the random coefficient, symmetrically fading in density along all dimensions of the modeled heterogeneity.

Now, let us consider a population which consists of three different (yet equally important) groups of agents, exhibiting preference heterogeneity in one regression parameter. In such a setting, the unimodal normal distribution cannot effectively accommodate population preferences, because the single-peakedness attribute will tend to inflate relative importance of the group which is closest to the population average, and suppress importance of the two groups which are further away from the average preferences. The resulting predictions

and policy recommendations will be hence inevitably biased towards the preferences of specific subgroup of the population.

Previous empirical and theoretical works (Apps & Rees, 2009) suggest that such multimodality might well be present in the context of female labor supply, where certain women are deeply engaged in their work, whereas others exhibit a strong preference to stay at home and look after children. Needless to say, these groups are likely to prove very different with respect to their time allocation preference, and hence the uni-modal model of unobserved heterogeneity might fail to capture this underlying dichotomy.

3.2 Alternative approach - latent class models

To circumvent the adverse implications of the parametric random coefficients model, researchers started to develop estimation procedures based on the multimodality assumption. These efforts led to a new type of models, called latent class models, which aim to identify subgroups of the population with unique (yet unobserved) characteristics, and estimate the original homogenous model for each of the latent subgroups.

A key advantage of the latent class approach is that it does not impose any distributional assumptions on the modeled class-level heterogeneity[2]. Regression parameters within the classes are allowed to vary freely, so that the method can be regarded nonparametric, or as Train (2008) puts it, *super-parametric*.

The latent class model in its general form can be represented as an extension of the multinomial logit presented in the Section 2.4. The individual conditional probability of choosing alternative j (see equation 2.5) becomes

$$\sum_{c=1}^{C} P(class_i = c) \cdot P(J = j|\mu, A_c, b_{ic}), \tag{3.1}$$

where $c = 1, ..., C$ is a finite set of latent classes, $P(class_i = c)$ is an agent i's unconditional probability of belonging to the given class c, and $P_i(J = j|\mu, A_c, b_{ic})$ is a conditional logit probability of observing the choice j, given a class-specific preference ordering (A_c, b_{ic}).

[2]Strictly speaking, by restricting ourselves to a finite number of latent classes, we are still imposing a distributional assumption on the population-level heterogeneity. However this assumption gets less restrictive, as the number of allowed classes gets larger.

The augmented individual conditional probability (3.1) hence accommodates C alternative values of the probability of choosing the observed outcome, which are weighted by the odds of being a member of the class corresponding to that value. The log-likelihood function is then derived out of the aggregated individual conditional probabilities,

$$\mathcal{L} = \sum_{i=1}^{n} \log \sum_{c=1}^{C} P(class_i = c) \cdot \prod_{j=1}^{J} P(J = j|\mu, A_c, b_{ic})^{d_{ij}} \tag{3.2}$$

The numerical maximization of the log likelihood is one of the potential downsides of the latent class models, because these can be very demanding in terms of computational complexity. Endeavors to refine the estimation procedure hence became one of the focal points of structural choice modeling and several alternative methods were developed to accommodate for the latent classes. These methods include the semi-parametric random coefficient mixed model developed by Heckman & Singer (1984) (including class shares as additional parameters in the classical ML optimization framework), the nonparametric expectation-maximization (EM) algorithm proposed by Train (2008) (utilizing class-share updating through sequentially optimized sub-models), and the nonparametric Bayesian estimation routine using Dirichlet Priors (Burda *et al.*, 2008).

We have decided to utilize the EM-algorithm procedure, as we deem the two latter approaches to be less fitting our analysis. In the case of the semi-parametric mixed models it is frequently observed that maximization algorithms fail to converge, because the maximized likelihood function turns overly non-smooth in the framework of class-specific utility parameters (Train, 2008). Regarding the nonparametric Bayesian estimation routine, (Burda *et al.*, 2008) affirm that the method proves efficient mostly due to the rich panel data structure of the underlying population sample. The data which we use for our analysis are however rather limited in this respect, so that we resort to the EM algorithm, which is described in the following section.

3.3 EM algorithm

The EM algorithm is a well-known econometric procedure, which has been extensively used since the seminal work of Dempster *et al.* (1977). It was in-

troduced as a remedy for missing data problems, but more recent works of Train (2008) or Pacifico (2009) prove that it is also well suited for identifying latent data classes in the framework of discrete choice models.

The algorithm identifies parameters and classes by evaluating relative goodness of fit of each class-specific parameterization when applied to the choices of agents observed in the population sample. That way, it can separate individuals whose decisions are more likely to be driven by different preference orderings.

A substantial benefit of the EM algorithm is that it always climbs uphill in the likelihood, eventually converging to a local maximum. In fact, the convergence is where the EM algorithm performs better than most of the gradient-driven estimators, such as semi-parametric mixed model. On the other hand, even the EM optimization can be difficult, as the local optima convergence requires practitioner to experiment with many starting values of the parameters, in order to reach the global optimum (Train, 2008). This is particularly important in empirical applications, where the actual latent classes are often loosely separated, blending together for agents with preferences on the margin of two different class parametrizations.

In the following description of EM algorithm, we will adopt the notation of Train (2008), so that the already established standards will change slightly.

The EM algorithm is an iterative method, which is based on recursive maximization of the following function,

$$\theta^{i+1} = \arg\max_{\theta} \sum_{n} \sum_{c} h_{nc}(\theta^{i}) \log s_c K_n(\beta_c). \tag{3.3}$$

where $K_n(\beta_c)$ represents the individual likelihood contribution, with β_c being a set of preference parameters attributed to class c,

$$K_n(\beta_c) = \prod_{j=1}^{J} P_n(J = j \,|\mu, A_c, b_{nc})^{d_{nj}}, \tag{3.4}$$

s_c is the aggregate class share, which is identical for all individuals in the sample, and which can be interpreted as an unconditional probability of belonging to the class c, with $\sum_{c=1}^{C} s_c = 1$.

And lastly, h_{nc} is the conditional posterior probability that individual n

belongs to the class c, which is computed using the two above established indicators, $K_n(\beta_c)$ and s_c.

$$h_{nc}(\theta) = \frac{s_c K_n(\beta_c)}{\sum\limits_{c'=1}^{C} s_{c'} K_n(\beta_{c'})} \quad , \quad \theta = \{\beta_c, s_c; c = 1, ..., C\} \tag{3.5}$$

The posterior probability reflects how does the model with class-specific parameterization β_c perform in explaining the agent n's choices, relative to the other classes. A high individual likelihood $K_n(\beta_c)$ of observing the selected choice as a member of class c is likely to be reflected in a high posterior probability of belonging to that class, as long as the corresponding unconditional class share s_c is high enough. That way, the EM algorithm evaluates individual goodness of fit of the alternative class models, and determines which agents are attributable to which latent class.

Relating to the previous section, we should note that the functional form (3.3) is different from the generalized latent class log-likelihood function as shown in equation (3.2). Nevertheless, Train (2009) proves that the two are interchangeable, with the maximized log-likelihood of function (3.2) being also attained by the EM recursion using function (3.3).

Now, having established the notation and the main components of our new functional form, we can describe the estimation process itself. Firstly, we are going to divide the functional form (3.3) into two parts.

Following Train (2008), the product $\log s_c K_n(\beta_c)$ can be rewritten as $\log s_c + \log K_n(\beta_c)$, hence splitting the optimization into two separate procedures,

$$\beta_c^{i+1} = \arg\max_{\beta_c} \sum_n h_{nc}(\theta^i) \log K_n(\beta_c) \tag{3.6}$$

$$s^{i+1} = \arg\max_{s} \sum_n \sum_c h_{nc}(\theta^i) \log s_c \tag{3.7}$$

with the second optimization procedure being actually a multinomial logit without other regressors than a constant, which proves to have a closed-form solution,

$$s_c^{i+1} = \frac{\sum\limits_n h_{nc}(\theta^i)}{\sum\limits_n \sum\limits_{c'} h_{nc'}(\theta^i)}. \tag{3.8}$$

This means that the updated aggregate class shares s_c^{i+1} can be computed straight from the posterior class-membership probabilities $h_{nc}(\theta^i)$.

Numerical optimization is hence required only for the procedure (3.6), with its optimized coefficients β_c^{i+1} and current class shares s_c^i being used to compute a new set of individual probabilities $h_{nc}(\theta^{i+1})$, as shown in the equation (3.5). The aggregate class shares s_c^{i+1} then get updated as a by-product of this numerical optimization through the equation (3.8).

The actual recursive procedure consists of the following steps:

1. Pick the starting values of parameters β_{c0} and class shares s_{c0}, and then split the sample population into C distinct subsamples (the subsamples act as initial draws for the latent classes).

2. For each subsample/class $1, ..., C$ estimate a separate multinomial logit model, as defined in Section 2.4.

3. Predict individual likelihoods for each class, and derive corresponding individual probabilities of class membership.

4. Derive new class shares based on individual probabilities.

5. Re-estimate the set of class-specific multinomial logit models, using class shares and membership probabilities as weights for individual likelihood contributions.

6. Repeat the steps 3-5 until reaching an appropriate convergence criterion[3].

One potential drawback of the EM algorithm is that the number of latent classes has to be chosen *a priori*, being fixed throughout the EM recursion. The selection of such a number is bound to be to some extent arbitrary, although we can use several identification strategies to make our decision more rigorous. For example, we can utilize our prior knowledge of the nature of classes within the population, or we can assess relative performance of the alternative models by comparing their likelihood-based information criteria, such as Schwarz-Bayes information criterion (BIC)[4].

[3]The convergence criteria of EM-algorithm can be based on the stability of regression coefficients, or on the sum of log-likelihoods (Train, 2008). In our case, we use the latter, requiring the difference between terminal likelihoods of the subsequent optimization rounds being smaller than $1.e^{-5}$ in five consecutive iterations.

[4]This approach is utilized in Train (2008) and Pacifico (2009)

Chapter 4

Data

4.1 Description of the dataset and sample selection criteria

The data we use in our analysis comes from the Household, Income and labor Dynamics in Australia (HILDA) Survey. This panel dataset contains rich socio-economic information on a representative sample (17,000 respondents) of the Australian population, which has been followed since year 2001.

Our models draw heavily on the data from its individual time-use survey[1], and also from its extensive child care questionnaire, which allows us to differentiate between utilization of formal and informal (unpaid) childcare.

We use four consecutive waves of the HILDA survey (2005-2008) which are turned into a pooled cross-section dataset. This pooling is necessary, as we are studying mothers with children in the pre-school age, who constitute only a small fraction of the whole HILDA sample, so that we would not have sufficient number of observations examining only a single wave[2].

[1]Time-use information are collected as a part of individual questionnaire, so that the person has to evaluate how she usually spends her time throughout the week, and fill in the hours accordingly. This approach is however less favorable than the "diary" collection mechanism where respondents are asked to record their daily activities as they occur in a real time, ensuring higher consistency than the one-off survey questions.

[2]This specific subsample is selected because Apps & Rees (2009) show that the mothers with preschoolers constitute a group whose labor participation is the most sensitive to the price and availability of child care. Also, focusing on preschool care is advantageous in that it can be considered as a rather uniform service, whereas the care for older children represents a very diverse mix of schooling, nursing, and other activities.

Apart from this baseline selection of mothers with preschoolers, we introduce a handful of other sample selection criteria to ensure consistency of our analysis. We exclude mothers if: they are single (neither married, nor cohabiting); at least one of the spouses is disabled, retired, or a full-time student; the husband is unemployed; the couple reports incomplete or implausible survey answers; or if they live in a multi-family household.

4.2 Description of the variables

After filtering out all the non-complying families, our final sample consists of 1465 couples. Their characteristics are summarized in the following table.

Table 4.1: Summary statistics, sample of couples with preschool children

Variable	Mean	Std. Dev.	Min.	Max.
Age - women	32.885	5.596	16	48
Age - men	34.931	6.276	18	58
Marital status (dummy)	0.826	0.379	0	1
Employment status - women (dummy)	0.555	0.497	0	1
Employment status - men (dummy)	1	0	1	1
Number of children, aged 0-4	1.398	0.559	1	4
Number of children, aged 5-9	0.437	0.676	0	4
Number of children, total	2.016	0.955	1	6
Paid work - women, weekly hours	13.929	15.782	0	80
Paid work - men, weekly hours	44.5	11.264	0	128
Housework - women, weekly hours	70.991	31.892	0	166.833
Housework - men, weekly hours	30.945	17.68	0	120
Informal childcare, weekly hours	23.335	18.046	0	140
Formal childcare, weekly hours	8.318	12.833	0	100
Formal childcare price, hourly	8.725	3.526	1.583	23.766
Annual wage - men	63373.581	35736.925	5136	357216
Annual wage - women	18514.129	23570.307	0	182256
Annual non-labour family income	6617.690	31778.24	0	683974
Number of observations		1465		

This table contains both the main variables entering our utility function, and the socio-demographic characteristics which enter the utility function as interaction terms. Furthermore, for comparison purposes, the gender-specific variables are presented also for males, even though this information is not directly used within our modeling framework.

A closer description of the variables is presented in the following subsections.

4.2.1 Basic characteristics and time-use indicators

As we can see, the average spouses are in their early thirties, with the man being about two years older than the woman. 83% of the couples are married, and the women tend to work in only 55% of the couples. The men are all working, which complies with the above imposed sample selection criteria. The unequal labor participation pattern is also reflected in the total hours worked, with a more than 30-hours gap between average values of the two genders, and very different distribution of hours, as shown in the Figure 4.1.

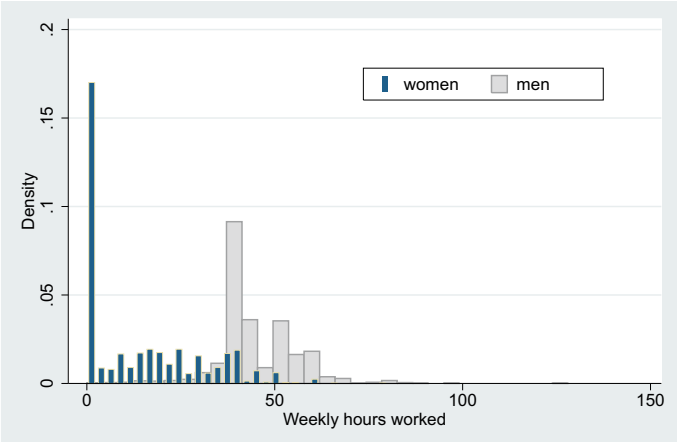

Figure 4.1: Distribution of weekly hours worked in families with preschool children

Apart from showing the non-participation spike in female distribution, the histograms also reveal that the men in most cases tend to work full-time (more than 35 hours a week[3]), whereas the women are often involved in some form of part-time occupation. These part-time contracts apparently exhibit considerable flexibility in working hours, with the distribution of working hours being relatively uniform within the part-time region.

We also see that there are several outliers who report unrealistically high hours of work, claiming to work more than 18 hours a day, seven days a week. To preserve consistency of our analysis, these observations are scaled down to comply with our basic "sustainability" criterion[4].

[3]The 35 hours is a threshold value of full-time occupation, as defined by Australian Bureau of Statistics (see `http://www.abs.gov.au/ausstats/abs@.nsf/Products/6103.0~Dec+2009~Chapter~Concepts%20and%20data%20items:%20F%20%E2%80%93%20K#FULL-TIME%20EMPLOYED`

[4]We assume that the sum of time spent on work and housework should not exceed 18 hours per day (=126 hours per week). If this threshold is surpassed, we decrease the two time allocations so that their sum equals 126 hours, and their relative size matches the unadjusted hours.

Relating to the observed dichotomy of working hours, it is not surprising that the intensity of work is reversed in the case of housework, as we can see from the average values of housework hours and their following histograms.

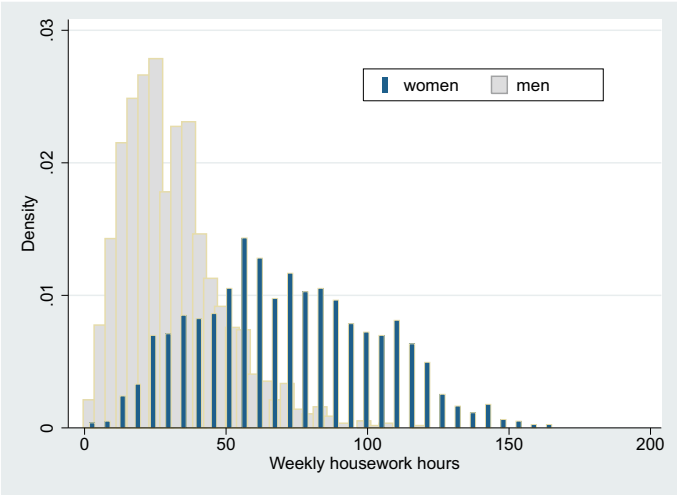

Figure 4.2: Distribution of weekly hours spent on housework in families with preschool children

It should be emphasized that the housework is considered here in a rather broad sense, comprising of all errands and chores related to the household. It also includes time spent with children, which can be interpreted as the childcare provided by one of the parents. That way, the housework indicator acts as a substitute for the childcare provided by others, be it formally, or informally.

To make the description of individual time-allocation variables complete, it remains to comment on the leisure.

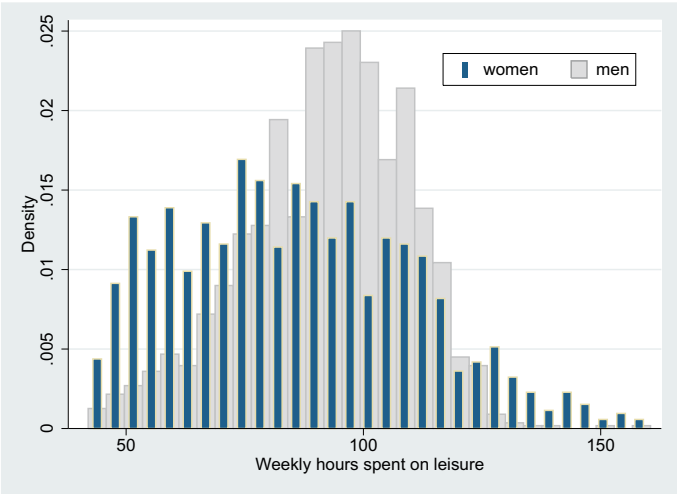

Figure 4.3: Distribution of weekly hours spent on leisure in families with preschool children

The leisure allocation is due to its construction bounded from above by 42 hours[5], and looking at the Figure 4.3, we can also see differences between the two sexes. On average, the men are having more leisure time than the women, and the distribution of their leisure hours is more spiked (which is a consequence of the relatively denser distributions of the two complementary allocations).

4.2.2 Childcare variables

In our analysis, we are differentiating between two types of childcare, the formal childcare provided by official institutions, such as kindergartens or care centers, and the informal childcare provided by friends, relatives, nannies, or by the husband.

The reasons for this separation are twofold. Firstly, the formal childcare differs from informal childcare in that it is recognized by the Australian fiscal authorities, so that its utilization makes the family eligible for reimbursement of considerable part of the corresponding costs.

The other reason is rather practical, and relates to the price of childcare. In the case of formal childcare, we can observe prices across all the families involved, whereas with the informal childcare, the service is often provided for free, or the price is just symbolic, not reflecting the actual value of time spent with children. Furthermore, the informal childcare does not fall under any official supervision, so that it suffers from the same problem as the childcare of school-aged children; the service accommodates many different ways of spending time with children, ranging from passive supervision, to specialized activities such as schooling. The lack of more detailed information about individual accounts of the informal childcare hence makes any efforts to impute corresponding market prices practically infeasible.

Therefore, within our discrete choice framework, we isolate the childcare decision to cover the formal care only, treating the informal care as exogenously given. The informal care indicator is then entering the utility function as one of the interaction terms, containing only hours used, without specifying the market price.

[5]We have already established that the leisure is regarded a remainder of the daily time endowment after subtracting work and housework activities. The 42-hours threshold is therefore a consequence of our sustainability criterion, as discussed above.

As for the descriptive statistics, the formal care is used by 43% of the families, whereas the utilization of informal childcare is almost universal (only 9 families report to not utilize any form of informal childcare). A distribution of the weekly hours of utilized childcare is presented in the Figure 4.4.

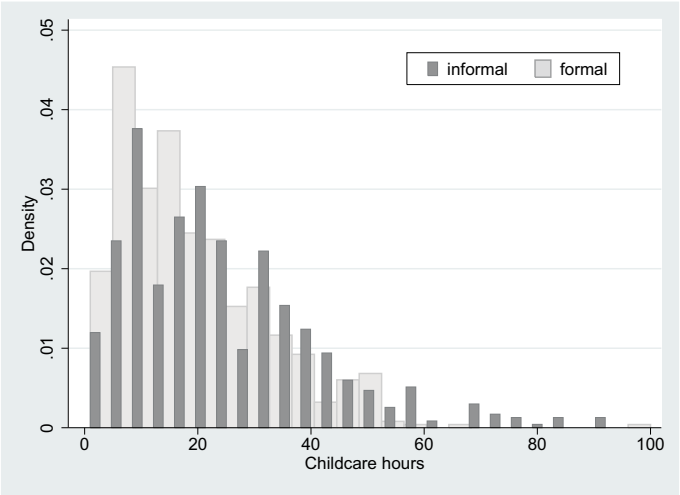

Figure 4.4: Distribution of weekly hours of informal and formal childcare, families with preschool children who are using childcare

We can see that the both profiles are relatively similar, although the formal care distribution does not go far above 60 hours per week. This would follow the rationale that the official institutions are closed during the weekends, so that the children are likely to spend this time with their family, or with an informal care-provider.

4.2.3 labor income, non-labor income

As we have outlined in the Section 2.1.2, the income indicator entering our utility function contains both labor and non-labor components.

The annual labor income for men and for women is derived from data on current weekly gross salary from all jobs, as reported in the HILDA questionnaire. The annual non-labor income is summed over the spouses, and includes reported business income[6], investment income, private domestic pensions and overseas pensions.

In the following figures, we present distributions of non-zero incomes for both subsets of family income. The zero-valued entries were omitted to render the rest of income distributions informative; in the case of female income

[6]Business income is excluded from the non-labor income in case that the given respondent runs an own business, or pursues another form of self-employment.

distribution, the zero-valued spike accounts for 45% of non-employed mothers. Similar mass-point is also observed in the case of non-labor income, representing 54% of families in the sample, for whom the labor wages are the only source of their income.

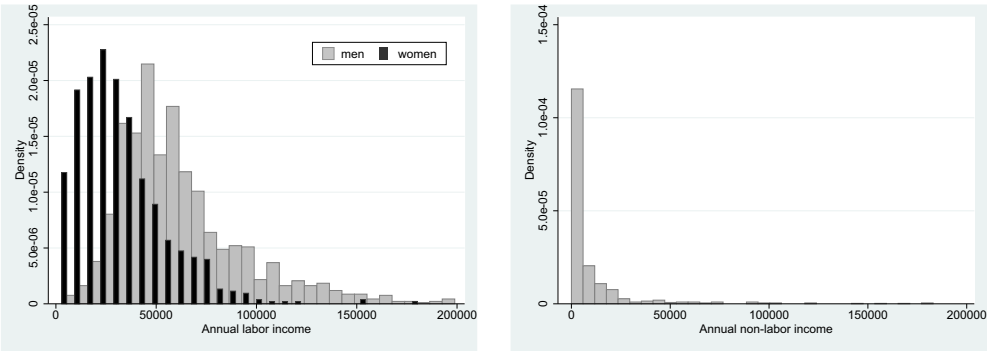

Figure 4.5: Distributions of labor and non-labor gross annual income, in Australian dollars, families with preschool children

From the income profiles we can see that the labor income distributions reflect gender disparities in the work intensity, with men earning on average more than women. As for the non-labor income, we observe that even in the subsample of families with non-zero incomes the distribution is skewed towards zero, with several outliers reporting larger gains from their businesses and investments.

As we have already discussed, these income indicators are used to derive the set of 64 alternative-specific family incomes. The gross incomes are plugged into the equation (2.1)[7], and the resulting gross family income is processed by our tax & benefit model. This leaves us with the net family income, which is deflated to the base year (2005) using the Australian consumer price index, and the resulting value is then used in the analysis.

A description of the tax and benefit model used in this calculation follows in the next section.

[7]With the mother's income being adjusted for her preferred work allocation.

4.3 Australian personal tax & social benefit system

The Australian personal tax system incorporates three main income tax measures, namely, Personal Income Tax (PIT), Low Income Tax Offset (LITO), and Medicare Levy (ML).

PIT is the cornerstone of Australian income taxation; it is a universal, individual tax with five progressive income brackets and marginal tax rate increasing from nil to 45% for the highest incomes[8].

The other two tax measures can be considered addenda to PIT. Firstly, LITO is a socially-targeted tax rebate for individuals who earn less than 30,000$ *p.a.* (this threshold is above the zero rate threshold of PIT). Below this threshold, tax payers receive tax offset of 750$ (or less, if they file tax proceeds lower than 750$). After surpassing 30,000$, the rebate is phased out by 4 cents in every dollar above the threshold, hence increasing the effective marginal tax rate on middle-class incomes by 4%.

The Medicare Levy is yet another tax measure which increases progression for the middle classes and relieves individuals with the lowest incomes. It is a 1.5% *joint*, flat-rate income tax which is payable after surpassing a specific low income threshold (around 16,000$ *p.a.* for single tax payers, and 30,000$ *p.a.* being a base threshold for couples, further increasing with the number of dependent children). Beyond the threshold, ML is phased in by 10 cents in every dollar, increasing the effective marginal tax rate of affected middle-class tax payers by 10%.

To illustrate the overall tax burden, we merge these three income taxes together and present resulting average and marginal individual tax profiles in the following graphs. The accompanying income densities reflect how is the tax burden spread among our subsample of mothers, and among the whole population sample.

[8]To illustrate the relative copmosition of the Australian tax system, we will restrict ourselves to tax rates and thresholds applicable in the fiscal year 2007-2008. The values corresponding to the period 2004-2007 do not systematically differ from those presented, and can be found on the website of Australian taxation office (`http://www.ato.gov.au/individuals/content.aspx?doc=/content/12333.htm`).

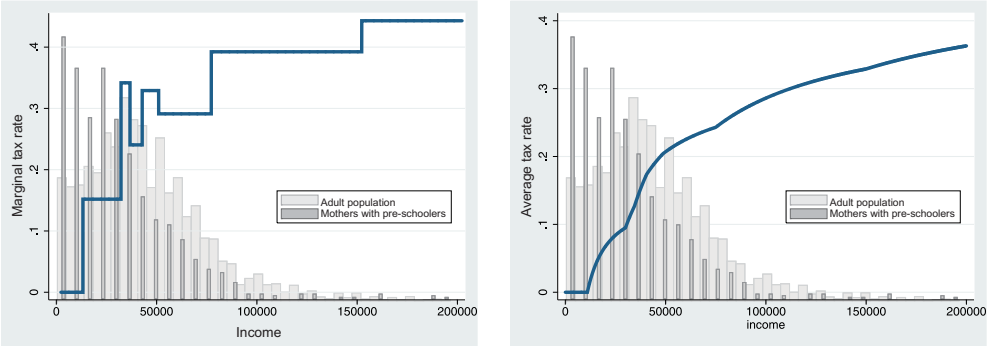

Figure 4.6: Individual income tax profiles, and annual income distributions

In the graph of marginal tax rate profile we can see two breaches of monotonic progression, which are caused by phasing out of the Low Income Offset, and phasing in of the Medicare Levy.

4.3.1 Social benefits

Due to the fact that we are explicitly interested in the childcare and its effects on decision making within households with children, it is essential to model all the family- and childcare-related benefit payments together with the income tax system. These payments include Family tax benefits A & B (FTB-A, FTB-B), Childcare Benefit (CCB), and Childcare Rebate (CCR).

Family Tax Benefit A is an income-tested transfer, aimed to help low income families with dependent children. The computation of effective FTB-A payment is somewhat involved, as it takes into account *joint* family income, number of children in the household and their age structure. The highest attainable benefit applies to families with school-aged children, where the maximum FTB-A rate attains 189$ fortnightly. After surpassing a family income threshold, this maximum rate is phased out in two tiers, firstly by 20 cents in every dollar to the base rate of 63$, and then by 30 cents in every dollar to nil.

Family Tax Benefit B is an additional social payment for families with the primary earner having annual income less than 150,000$, and secondary earner being in the household, or earning less than 4,380$ *p.a.*. The effective rate of FTB-B is adjusted for the age of the youngest child (125.02 for pre-schoolers, and 87.08 for schoolers), but it does not increase with more children in the household. After surpassing the second earner's threshold, FTB-B is phased out by 30 cents in every dollar of excess income.

As we can see, both family tax benefits can be interpreted as income tax shifters, changing the marginal and average tax rate through their eligibility and phasing-out conditions. We can hence augment the previously depicted marginal and average tax rates and show the pattern of effective income tax, utilizing all the discussed fiscal measures.

In the following graphs, we show the tax profiles for two separate households, the first being a household with a single earner, and the second where both spouses are employed and earning the same wages (so that the joint family income is doubled in the case of two-earner household).

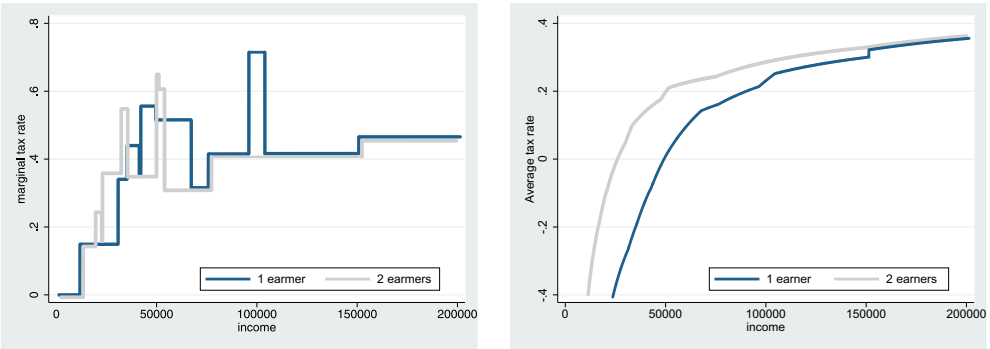

Figure 4.7: Individual income tax profiles for one-earner and two-earner households, including income tax measures and family tax benefits

The graphs show that the introduction of Family tax benefits turns the original income tax progression overly non-monotonic, with the marginal tax rate exhibiting inverse U-shaped profiles among the middle-income families. As a result, the tax burden gets disproportionally shifted towards the average earners, leaving the high-tier tax rates unchanged.

This disproportionate burden has been recognized and thoroughly criticized by Apps & Rees (2009), who point out an another alarming characteristic of the FTB system - the deterrent effect on female labor force participation.

The deterrent effect arises through phasing-out of the both benefits. Recalling that the FTB-A is tested on joint income of the spouses, the middle-class mothers who start working as a secondary earner are likely to be subject to phasing-out of the FTB-A benefit, irrespective of their earnings. This translates into an increase of marginal tax rate by either 20% or 30%. Furthermore, the phasing-out of FTB-B, which is derived explicitly from the income of sec-

ond earners can raise the MTR by additional 30% even for symbolic secondary incomes. As a result, in certain middle-class families, part-time working mothers earning around 10000$ *p.a.* can be facing marginal tax rates as high as 77%. For that reason, many mothers choose to stay out of the labor force, and specialize in the houshold goods production.

The last two transfers entering our analysis are Childcare Benefit (CCB) and Childcare Rebate (CCR), which constitute supplementary family payments, directly targeted to facilitate the use of formal (paid) childcare.

CCB is probably the most complex social payment in our model, as it depends (among other things) on age structure of children, number of children, type of childcare used, and the hours of childcare used. Furthermore, the phasing-out system is also dependent on the age characteristics of children, making its accurate description in this chapter practically infeasible[9].

Equally important as CCB is the Childcare Rebate, which reimburses families for their claimed childcare expenses. This reimbursement is very generous, as it can amount up to 50% of the net childcare expenses (that is, after subtracting CCB). The CCR rate is not income-tested, but it has got an upper cap on the amount of expenses which can be reimbursed. For year 2008, this cap was 4,354$ *p.a.*.

[9]All the CCB guidelines, historical thresholds, and examples of the rate calculation, can be found on the governmental website, `http://www.fahcsia.gov.au/guides_acts/fag/faguide-3/faguide-3.5.html`.

Chapter 5

Results

In this section we are going to analyze the outcomes of our regression analyses, both in the baseline homogenous specification, and also in the extended form with the latent classes modeled through the EM algorithm. A discussion of the baseline specification follows.

5.1 Baseline model, no heterogeneity

Our baseline model is a direct implementation of multinomial logit, as outlined in the Section 2.4. The main disadvantage of such model is that it is not able to control for unobserved heterogeneity. On the other hand, its computation time is minimal, and if the homogenous assumption proves valid, the results are consistent. The corresponding regression output is shown in first column of Table 5.1.

Before we comment on individual regression coefficients, it is convenient to recall that our utility function (2.2) is built around the four time-use and income indicators, and that it can be split into two separable components: the matrix of quadratic coefficients A, and the vector of linear parameters and socio-demographic interactions b.

For ease of interpretation, we start with the discussion of interaction terms which are included below the divider of Table 5.1. The coefficients of interaction terms can be interpreted as shifters of the effect born by the linear coefficients of time-use. Depending on their value and on the corresponding socio-demographic characteristic, they can adjust size, or even change a sign of these linear effects. To illustrate, from the results we can see that the provi-

Table 5.1: Regression results-baseline model, and model with 2 lat. classes

	no heterogeneity	2-class (1)	2-class(2)
income	-.422	-1.320	.540
	(.135)***	(.301)***	(.270)**
paidcare	1.949	19.226	-15.112
	(5.922)	(12.094)	(10.059)
housework	7.785	25.423	.185
	(4.645)*	(10.632)**	(8.964)
leisure	13.837	35.164	-1.534
	(4.883)***	(10.982)***	(11.371)
paidcare^2	-.069	-.242	.053
	(.011)***	(.027)***	(.019)***
housework^2	.026	-.180	.192
	(.009)***	(.024)***	(.027)***
leisure^2	-.058	-.273	.067
	(.010)***	(.026)***	(.030)**
inc*paidcare	.008	.015	.011
	(.003)**	(.006)**	(.005)**
inc*housework	.011	.028	.015
	(.003)***	(.007)***	(.006)**
income*leisure	.011	.028	.017
	(.003)***	(.007)***	(.007)**
paidcare*housework	-.090	-.138	-.050
	(.005)***	(.016)***	(.009)***
paidcare*leisure	-.090	-.144	-.039
	(.006)***	(.017)***	(.010)***
housework*leisure	-.014	-.228	.151
	(.009)	(.024)***	(.025)***
paidcare*log(age)	.724	-8.331	9.623
	(3.452)	(6.915)	(5.899)
paidcare*log(age)^2	-.064	1.318	-1.390
	(.502)	(.995)	(.862)
paidcare*married	-.045	.004	-.027
	(.052)	(.114)	(.079)
paidcare*No. dependent children	-.091	-.242	-.079
	(.035)***	(.084)***	(.050)
paidcare*children aged 0-4	.233	.498	.213
	(.054)***	(.103)***	(.089)**
paidcare*children aged 5-9	.063	.354	-.105
	(.046)	(.096)***	(.082)
paidcare*informal care	-.008	-.008	-.008
	(.001)***	(.002)***	(.002)***
housework*log(age)	-4.295	-8.464	-4.404
	(2.719)	(6.128)	(5.270)
housework*log(age)^2	.609	1.131	.672
	(.399)	(.892)	(.777)
housework*married	-.015	-.087	.054
	(.043)	(.101)	(.079)
housework*No. dependent children	-.089	.212	-.287
	(.030)***	(.087)**	(.092)***
housework*children aged 0-4	.387	.177	.673
	(.044)***	(.111)	(.111)***
housework*children aged 5-9	.143	-.163	.379
	(.038)***	(.103)	(.103)***
housework*informal care	-.011	-.022	-.005
	(.0009)***	(.002)***	(.002)***
leisure*log(age)	-6.416	-12.669	-1.248
	(2.869)**	(6.326)**	(6.702)
leisure*log(age)^2	.869	1.701	.124
	(.422)**	(.923)*	(.991)
leisure*married	-.027	-.110	.078
	(.046)	(.099)	(.106)
leisure*No. dependent children	-.019	.272	-.195
	(.032)	(.095)***	(.108)*
leisure*children aged 0-4	.271	.104	.485
	(.047)***	(.116)	(.135)***
leisure*children aged 5-9	.108	-.172	.301
	(.043)**	(.109)	(.132)**
leisure*informal care	-.012	-.025	-.005
	(.001)***	(.002)***	(.002)**
n	1465	1465	1465
population share	100%	67.573%	32.427%

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

sion of informal childcare reduces the utility derived from the formal childcare, suggesting that the two care types act as substitutes, rather than complements.

Another interesting effect is the one of age; the quadratic age interactions in the table of results show that the formal childcare is more valuable for younger mothers, whereas the leisure and housework get more appreciated as the mothers become older.

However, despite the straightforward interpretation, the interaction terms constitute a mere fraction of the compound effect of the time-use indicators, which are at the focal point of our analysis. These effects are also much harder to evaluate based on isolated regression coefficients, because a single change of any of the main regressors will distribute itself through both the quadratic matrix A and the linear vector b, hence obscuring the importance of partial effects associated with the individual coefficients.

Therefore, to evaluate how such variables affect individual utility, we have to aggregate all the coefficients and interaction terms corresponding to a given regressor into one composite indicator. A natural candidate for this kind of indicator is marginal utility, measuring total change of utility associated with an infinitesimal alteration of the regressor, *ceteris paribus*.

$$MU(t_w^{hw}) = \frac{\partial V\left(\mu(t_w^{hw}, t_w^l, pc, y)\right)}{\partial t_w^{hw}} \tag{5.1}$$

The marginal utility is individual-specific, because the utility function V contains both socio-demographic characteristics and individually chosen time-allocations. Therefore, to comment on the overall effects of the time-use and income variables, we derive two set of aggregate indicators: average marginal utilities, and shares of individuals exhibiting negative marginal utility from the given variable. Their values are presented in the following table.

marginal utility:	full sample avg.	childcare users avg.	negative fraction
Income	0.0873	0.0933	0.0444
Paid care	-0.4007	-0.4381	0.8956
Housework	0.4305	0.0852	0.1884
Leisure	0.2925	0.0283	0.2375

Table 5.2: Average marginal utility of the main regressors and fraction of the population with negative marginal utilities, homogenous model

As we can see, the marginal utilities are on average positive for income, housework and leisure, and negative for formal childcare. The share of mothers exhibiting disutility from additional paid care is very high, reaching to almost 90% of the sample. The latter negative shares are substantially lower, attaining mere 4% for income, and about 20% for both housework and leisure.

The adverse marginal effect of the paid care is counterintuitive, especially as it is observed also within the subsample of mothers who are actually using paid care (see second column of the Table 5.2). This implies that even the active users would be better off with reduced or no formal childcare, compared to their current allocation.

Nevertheless, such fallible outcome can be potentially explained by unobserved heterogeneity. An important observation in this respect is that our sample contains a large fraction of mothers who are using no formal childcare at all (almost 60% of all observations). This constitutes a disproportionate spike in the distribution of childcare hours, especially compared to the subsample of active users (see Figure 4.4).

Now, if the childcare allocation decisions are to some extent driven by the unobserved heterogeneity, then our baseline multinomial logit is likely to assign disproportionate weight to the latent preferences of mothers who display strong distaste for formal childcare. This will rebound in biased coefficients of childcare, 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.

5.2 Allowing for unobserved heterogeneity through the EM-algorithm

The latent class augmentation brings another layer of complexity into the modeling process. This reflects in increased processing time, however the computation burden is more than compensated by the increased precision and consistency of our regression estimates.

Firstly, for the sake of clarity and ease of interpretation, we will focus on the model with two latent classes, then we will extend our analysis to more stratified models, identifying the model with optimal number of classes.

5.2.1 Model with two latent classes

The results of the two-class model can be found in the second and third column of Table 5.1. Looking at the estimated coefficients, we observe that the preference ordering display substantial differences between the two latent classes, with the first class parameters proving valid for approximately 68% of the mothers, and the latter fitting the remaining 32%.

Corresponding marginal utilities for both classes are presented in Table 5.3, and although their levels are mutually incomparable due to the implicit logit normalization, their signs still bear interesting information for our analysis.

marginal utility:	full sample	childcare users	negative fraction
Class 1			
Income	-0.0100	-0.0094	0.5167
Paid care	-0.2792	-0.6528	0.6614
Housework	0.3367	0.1715	0.2403
Leisure	0.1727	0.0628	0.3570
Class 2			
Income	1.2661	1.6146	0.0000
Paid care	-0.0792	0.3485	0.7952
Housework	0.4405	-0.2246	0.1959
Leisure	0.3064	-0.2307	0.2423

Table 5.3: Marginal utility of the main regressors and fraction of the population with negative marginal utilities, 2-class model

On average, the first class exhibits negative marginal utilities for both income and formal childcare, whereas the second class attains the negative sign for childcare only. The observation of the distaste for additional income raises the same questions as the distaste for formal childcare, which we have discussed in the previous section. And also in this case, the reasons for such finding relate to sample composition.

Regarding the composition of our first class, one socio-demographic characteristic stands out - it is more populated by mothers who are either in the household, or working part-time hours[1]. Their dominance in the subsample causes that in our time-allocation framework, many class-1 agents choose the alternatives with zero hours of paid work, and hence also with the lowest incomes in their choice set. Implicitly, the multinomial model then translates this behavior as a preference for lower incomes, and the additional income can indeed end up exhibiting negative marginal utility.

[1]The first class accommodates 78% of mothers who are not employed, or working less then 25 hours per week.

This counterintuitive finding is associated with the fact that the time spent on housework does not have direct pecuniary valuation. As Apps & Rees (2009) claim, the spouses who stay in the household can be hardly considered inactive - they are still likely to create a value added for the family, doing tasks and chores which would have to be outsourced and paid for otherwise. In this context, the observed preference for lower incomes does not mean that the mothers aim for less family wealth, but that they are working in a way that is not financially appraised by the market.

The average marginal utility of formal childcare retains negative sign in both parameterizations, which might suggest that the stratification into two latent classes is enough to overcome the biases caused by the unobserved heterogeneity. Nevertheless, looking at the marginal utilities of active childcare users, we can see that for the second class the sign of childcare indicator proves positive. Such identification is promising, as it supports observed utilization of childcare among the active users.

This particular change of sign also relates to the composition of the classes, with the second class being more populated by mothers who are using formal childcare[2]. Therefore, their preferences are likely to play greater role in the optimization, resulting in the positive marginal utility of childcare for the subsample of its users.

However, despite the observed differences in the group composition, it would be futile to conclude that the two-class EM model has effectively separated preferences of housewives on one side, and working mothers on the other side. Such identification is unlikely even in the models with many latent classes, because the underlying individual heterogeneity (and therefore also the resulting class composition) is spread along various dimensions of the data, with the decision whether to work or not being only one of them.

The latent class approach is trying to separate subgroups of the population with relatively homogenous preferences, but given that the number of classes is always to some extent arbitrary and the unobserved heterogeneity is a continuous multi-dimensional attribute of the data, we are unlikely to get the degree of separation which would directly match any observable characteristic.

[2] 43% of mothers in the second class are active childcare users, whereas in the first class, the corresponding share is 37%.

5.2.2 Identification of the optimal number of classes

As discussed in the Section 3, a key step of the EM algorithmic estimation is the selection of the number of latent classes. This decision is very important; on one hand, the higher is the allowed number of heterogeneous groups, the higher is the asymptotic precision of our model. However, on the other hand, more stratified models require much more data, as we estimate new set of regression coefficients with each class introduced into the optimization procedure. The determination of the number of classes is therefore crucial for identifying the optimal model.

Following (Train, 2008), we have compared our models on the basis of their Schwarz-Bayesian information criteria (BIC),

$$BIC \approx -2 \cdot \log(L) + k \cdot \log(n), \tag{5.2}$$

where L is the likelihood (as in equation (3.3)), k is the number of free parameters in the model and n is the number of observations in our data sample. The multiple-class models yielded following statistics:

No. of classes	Log-likelihood	BIC
1	-4695.31	9645.74
2	-4144.57	8813.99
3	-3829.09	8445.47
4	-3578.40	8206.52
5	-3307.73	7927.59
6	-2789.46	7153.49
7	-2649.79	7136.57
8	**-2461.64**	**7022.70**
9	-2343.63	7049.10

Table 5.4: Bayesian Information Criteria for multi-class models

As we can see, the 8-class model attained lowest BIC, and should be therefore used as the most reliable model for our analysis.

Regarding the results for multiple-class models, we refrain from presenting the regression output for each new model, as these are not very informative *per se*. Rather than that, we will show how do the differently stratified models perform in the policy-reform simulations which are presented in the following section.

5.3 Policy simulations

The ultimate goal of labor supply analyses is to predict how people respond to changes within the given economic setting. These predictions are performed by means of simulations, and in this paper we are evaluating following policy changes: firstly, we simulate two basic adjustments of the aggregate price level, a 10% increase of gross wages of mothers, and a 10% increase of gross prices of formal childcare. Then, we pursue a policy simulation in the spirit of Apps & Rees (2009), building on their critique of FTB and other joint-income fiscal measures (see Section 4.3.1). We propose an alternative system of taxes and benefits which aims to be less distortive to the female labour supply than the current one, and we estimate its impact on decision making within the analyzed households.

5.3.1 Changing the gross wages and gross childcare prices

Before we start evaluating the results of our two simulations, we should discuss how can we interpret such policy changes in the context of Australian fiscal system, and how can we measure their impact on the society.

In the case of gross wage change, the increased wages will reflect in higher disposable incomes of working mothers, so that the first simulation can be interpreted as a universal reduction of the personal income tax. However, this does not mean that the effective marginal tax rate is going to drop uniformly among all families in the sample; by increasing the gross incomes, some mothers are going to face changes in their effective tax & benefit rates (as discussed in Section 4.3), so that the resulting tax change will prove disproportionate among different groups of families.

And similar interpretation also applies for the second policy change - the increase in the gross price of childcare can be considered an uneven tax on the net formal childcare costs, depending on the effective CCR & CCB rules.

The impact of such reforms is measured in terms of aggregate elasticities, that is, ratios of percentage changes in the analyzed variables (in our case, time-use allocations), and percentage changes in the underlying policy instruments (in our case, wages and prices). These indicators display how pronounced are changes in the post-reform allocations, relative to the magnitude of policy

change itself.

Computation of the aggregate elasticities is executed in the following way: firstly, we derive the pre-reform allocations by averaging individual choice probabilities predicted by our logit model, and multiplying them by the corresponding choice-specific time-use allocations. This will provide us with a set of pre-reform allocations, which is in fact a simulation of average time-use hours, based on the optimized parametrization of our model[3].

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 individual choice probabilities we substitute the original dataset by an augmented version, where the incomes were altered to reflect the assessed reform. By feeding the augmented data into our optimized 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 elasticities is a matter of straightforward arithmetics. The results for models with different composition of classes are provided in Table 5.5.

Gross wage increased by 10%

No. of classes	1	2	3	4	5	6	7	8	9
Formal care hrs.	0.03	0.19	0.21	0.27	0.16	0.17	0.19	0.16	0.17
	(0.013)**	(0.042)***	(0.048)***	(0.06) ***	(0.055)***	(0.042)***	(0.061)***	(0.038)***	(0.052)***
Work hrs.	0.01	0.31	0.32	0.24	0.26	0.21	0.20	0.18	0.14
	(0.018)	(0.054)***	(0.046)***	(0.061)***	(0.058)***	(0.05)***	(0.068)***	(0.076)**	(0.091)
Housework hrs.	0	-0.05	-0.06	-0.03	-0.02	-0.01	0.02	-0.01	-0.01
	(0.004)	(0.009)***	(0.001)***	(0.011)***	(0.009)**	(0.011)	(0.012)	(0.008)	(0.014)

Gross childcare price increased by 10%

No. of classes	1	2	3	4	5	6	7	8	9
Formal care hrs.	-0.05	-0.42	-0.42	-0.49	-0.48	-0.46	-0.38	-0.48	-0.40
	(0.016)***	(0.056)***	(0.059)***	(0.065)***	(0.084)***	(0.059)***	(0.123)***	(0.09)***	(0.185)**
Work hrs.	-0.01	-0.10	-0.10	-0.16	-0.10	-0.10	-0.12	-0.10	-0.10
	(0.006)*	(0.019)***	(0.02)***	(0.029)***	(0.026)***	(0.024)***	(0.058)**	(0.022)***	(0.219)
Housework hrs.	0	0.02	0.03	0.02	0.01	0	0	0.01	0
	(0.002)	(0.005)***	(0.005)***	(0.007)**	(0.006)*	(0.007)	(0.006)	(0.017)	(0.037)

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

Table 5.5: Elasticities of the time-use allocations with respect to the changes in gross wages and gross childcare prices

[3]In ideal case, these simulations should be close to their empirically observed counterparts. As for the 8-class model, the simulated daily averages are 1.42 hours for formal childcare, 2.06 hours for work, and 9.07 hours for housework. The corresponding observed mean values are 1.2, 2.01 and 10.23, respectively, with the mean value of housework being slightly larger due to the effect of outliers reporting excessive hours.

Looking at the figures, the most striking difference is when we compare the homogenous model with the models incorporating unobserved heterogeneity. When we allow for latent classes, the predicted responses to policy reforms increase by an order, remaining relatively steady among different class-parameterizations. Such dichotomy of our results reflects how important is to control for the unobserved heterogeneity in the framework of female labor supply. Furthermore, the introduction of latent classes raises significance of the aggregate elasticities, even though we observe it fading again as we allow for further classes [4].

Relating to the BIC selection procedure in Section 5.2.2, following inspection of actual coefficient values will focus on the optimal parametrization, that is, the 8-class model.

In the case of the gross wage increase, we observe time-use shifts which go hand in hand with the common sense. With the reform in place, it is more profitable to enter the job market, and so the mothers start to work more; 10% wage increase reflects in 1.8% rise of average working hours. However, their children still need to be looked after, and hence the mothers increase consumption of the bought-in childcare, which rises by 1.6% of hours demanded.

The housework elasticity proves to be negative and insignificant, which has got several implications. Firstly, the negative sign implies that higher wages make the mothers work less in the household. However, the insignificance of the coefficient suggests that apart from housework, mothers are also likely to give up their leisure activities in order to work more. This claim is further supported by the observation that the -0.1% shift of housework hours is not enough to counterweight the reallocation of time towards paid work.

Turning to the impact of the rise in childcare prices, it is not surprising that the most responsive elasticity is its own. With 10% rise in the childcare prices, the demand for formal childcare falls by 4.6%. This in turn causes the mothers to work less, as they have to substitute for the forgone bought-in services. Interestingly, the change of housework hours is insignificant and

[4] Standard errors on the elasticities were computed through 100 Monte Carlo simulations, re-estimating the impact of the reforms under alternative sets of preference coefficients. 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 parametrizations.

closing to zero, so that by working less, the mothers are implicitly increasing their consumption of leisure. One possible explanation is that the extra time spent with children is regarded as an additional leisure. Alternatively, there might have been some reallocations of chores among the spouses, resulting in more free time for the mothers.

5.3.2 Alternative system of taxes and benefits

As we have already discussed in Section 4.3.1, the joint-income fiscal measures, such as FTB or Medicare Levy, are likely to pose a substantial drawback for female labor supply. For that reason, we want to test whether an alternative tax scheme could prove less distortional, potentially improving the labor conditions of employment-seeking mothers.

The pursued policy simulation relates to the work of Apps & Rees (2009), who propose a fiscal system based on individual income taxation and universal benefits, so that all excessive joint-income taxes and phasing-out schemes are either discarded, or absorbed by the baseline progressive individual income tax.

To accommodate for such policy recommendation, we discard the ML surcharge and turn the FTB benefits independent on the family income. More specifically, the FTB payments retain their original stratification based on given family characteristics (See Section 4.3.1), but we do not allow for income-tested phasing-out of the payments. As a result, all families will receive full amount of the FTB benefit, irrespective of their wealth.

These two policy changes are however bound to reflect themselves in a substantial increase of government spending. Following Apps & Rees (2009), we compensate for the surge of benefit payments by a proportional increase of individual income taxes (represented by PIT & LITO).

Assuming that the individual time allocations remain unchanged, a full compensation of the reform-induced spending requires the income tax proceeds to increase by 26.76%[5]. Therefore we multiply each rate of the income tax by 1.2676, making the resulting personal income tax scheme more progressive than the original.

[5]The maintained assumption of no time allocation changes is bound to be disproved by the simulation results, but it serves as a useful reference point for quantifying the *ex-ante* compensating tax change. Under this assumption, the new system is modeled to be equally efficient as the original, yielding the same tax proceeds. This equality is convenient, because by predicting the actual reform-induced changes of working hours and childcare prices, we are able to quantify additional gains (or losses) to the aggregate tax proceeds, hence evaluating relative efficiency of the new system, compared to the original fiscal setting.

The Figure 8 presents differences in the net tax positions of our sampled families, induced by the proposed reform. The differentials are ordered by the corresponding pre-reform net incomes, so that we can see how the tax burden shifts over different income groups[6].

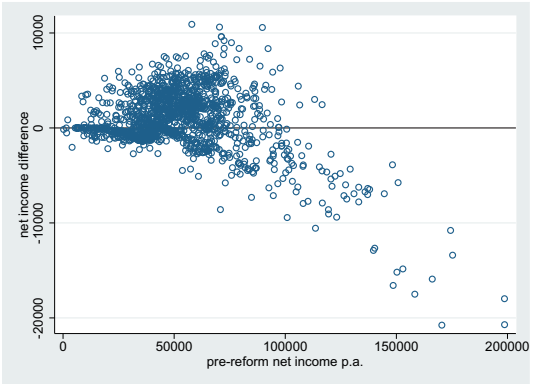

Figure 5.1: Post-reform differences in the net tax positions of the families, ordered by their pre-reform net incomes.

Looking at the scatterplot, we see that the increased progression of the post-reform tax system moves the tax burden back towards the higher incomes, leaving the middle-class families better off, compared to their original net tax positions.

It is also interesting to note that the increased progression can be considered a remedy for the forgone phasing-out of the family tax benefit, because the universal FTB payments will get eventually subtracted from the incomes of high-earning families through the effect of increased income taxes.

No. of classes	1	2	3	4	5	6	7	8
Formal care hrs.	0.77%	2.5%	2.57%	4.10%	2.61%	2.28%	1.84%	2.34%
	(0.133)***	(0.528)***	(0.577)***	(0.95)***	(0.796)***	(0.602)***	(0.867)**	(0.663)***
Work hrs.	0.73%	3.94%	4.13%	3.73%	3.96%	3.37%	2.01%	2.79%
	(0.253)***	(0.767)***	(0.643)***	(0.98) ***	(0.973)***	(0.799)***	(1.024)**	(1.11)**
Housework hrs.	-0.1%	-0.67%	-0.71%	-0.45%	-0.38%	-0.15%	0.07%	-0.18%
	(0.005)**	(0.124)***	(0.125)***	(0.165)***	(0.148)**	(0.141)	(0.175)	(0.138)

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

Table 5.6: Percentual changes in time allocations after the FTB reform

As for the behavioral implications of our simulated policy change, Table 5.6 presents changes in the time allocations induced by the reform.

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

Similarly to the previous simulations, we observe large discrepancy between size of the changes predicted by the homogenous and latent-class models, with the results of latent-class models proving relatively stable among different specifications.

Regarding the coefficient values, we will again restrict ourselves to the 8-class specification, in which we observe a 3.44% increase in the hours of work, a 3.24% increase of hours of formal care, and a 0.15% decrease in the hours of housework, although the latter change does not prove to be significant. Therefore, we can confirm that the proposed policy change is able to promote employment decisions among the population of mothers with preschool children.

Furthermore, we analyze the net fiscal effect of the reform by comparing the tax proceeds of newly employed mothers, and benefit payments for increased utilization of formal care. According to our calculations, the ratio between additional post-reform proceeds and costs is predicted to be approximately 3:2, so that the reformed fiscal system may well prove more efficient than the current one.

Chapter 6

Conclusion

In this thesis we have analyzed preferences of married and cohabiting mothers with pre-school children in the framework of their labor supply choices, and other time allocations. We have focused our analysis on identification and understanding of the unobserved heterogeneity, which has proven to play a dominant role for the decision making of mothers in our data sample.

Regarding the importance of unobserved heterogeneity, we have observed that a significant fraction of mothers in the household exhibits strong distaste for formal childcare, implying that they would prefer to raise their children by themselves despite potential pecuniary gains from engagement in a paid work. On the other hand, we have also observed a subset of mothers who are working, and who are actively using formal childcare, which they consider a beneficial service on its own. The apparent dichotomy of unobserved tastes among the identified latent groups has turned the homogenous model with no unobserved heterogeneity clearly invalid. Its optimized parameters failed to capture the actual preferences underlying individual decision making, hence distorting the policy simulations which were based on the baseline model specification.

To control for unobserved heterogeneity, we have estimated a series of latent-class multinomial logit models, and chose the 8-class model to be the best parametrization. This model was identified to perform optimally, balancing goodness of fit on one side, and parsimony on the other side of the chosen selection criterion. Then, to assess responsiveness of our sample to changes in the fiscal system and price levels, we have conducted a series of policy reform simulations, increasing gross wages of mothers and gross childcare prices in the first two reforms, and altering the joint-income fiscal measures in the third reform.

The first two policy simulations have proven that the mothers are responsive both to the shifts of wages and childcare prices. They tend to work more and consume additional childcare when it is more profitable to do so, and they are likely to stay in the household and look after the children when the childcare prices or labor conditions turn adverse. This identification shows that paid work and formal childcare act as natural complements, and a change of either component is bound to have the same effect on its counterpart.

Furthermore, the results have shown that important shifts of labor supply and childcare demand can remain unidentified when the unobserved heterogeneity is left untreated, underestimating the size of predicted changes in the women's behavior by an order.

As for the third simulation, we have discussed that the joint-income tax measures are likely to prove adverse for the labor supply decisions of women. Through the simulation, we have shown that the fiscal system can be made more favorable for employed mothers by discarding any distortive tax measures, and incorporating them into the individual income tax. In such setting, the mothers are predicted to shift their time allocations towards market work, and also to utilize more formal child care. Furthermore, the results also show that the resulting tax system may prove more efficient on the aggregate level, extracting additional tax proceeds from the incomes of newly employed mothers.

It remains to be emphasized that our current model specification is by no means terminal, and would benefit from a handful of extensions; firstly, our analysis would benefit from exploiting the panel-data structure of HILDA dataset, controlling for time-stable individual effects. And secondly, although we consider the current method of treating unobserved heterogeneity to perform well, it could be worthwhile to assess stability of our results by using alternative techniques of controlling for unobserved heterogeneity, such as random coefficient mixed logit model, or approaches utilizing Bayesian nonparametric methods.

Bibliography

Apps, P. & R. Rees (2009): *Public Economics and the Household*. Cambridge University Press

Berndt, E., B. Hall & R. Hall (1974): "Estimation and inference in non-linear structural models." In "Annals of Economic and Social Measurement, Volume 3, number 4," NBER Chapters, pp. 103–116. National Bureau of Economic Research, Inc

Bhat, C. (2000): "Flexible model structures for discrete choice analysis." In "Handbook of transport modelling," volume 1, pp. 71–90. Pergamon, Amsterdam, NL

Blundell, R. & T. MaCurdy (1999): "Labor supply: A review of alternative approaches." In "Handbook of Labor Economics," volume 3 of *Handbook of Labor Economics*, chapter 27, pp. 1559–1695. Elsevier

Blundell, R. & A. Shephard (2008): "Employment, hours of work and the optimal taxation of low income families." *IFS Working Papers* **(W08/01)**

Burda, M., M. Harding & J. Hausman (2008): "A Bayesian mixed logit & probit model for multinomial choice." *Journal of Econometrics* **147(2)**: pp. 232–246

Cameron, A. C. & P. K. Trivedi (2005): *Microeconometrics*. Cambridge University Press

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

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

Heckman, J. (1979): "Sample selection bias as a specification error." *Econometrica* **129(1)**: pp. 153–161

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

Keane, M. & R. Moffitt (1998): "A structural model of multiple welfare program participation and labor supply." *International Economic Review* **39(3)**: pp. 553–589

Pacifico, D. (2009): "On the role of unobserved preference heterogeneity in discrete choice models of labor supply." *Working Paper Univ. Modena* **1(1995)**: pp. 1–24

Ruud, P. (1991): "Extensions of estimation methods using the EM algorithm." *Journal of Econometrics* **49(3)**: pp. 305–341

Train, K. (2008): "EM algorithms for nonparametric estimation of mixing distributions." *Journal of Choice Modelling* **1(1)**: pp. 40–69

Train, K. (2009): *Discrete Choice Methods with Simulation.* Cambridge University Press

Van Soest, A. (1995): "Structural models of family labor supply: A discrete choice approach." *Journal of Human Resources* **30(1)**: pp. 63–88

Van Soest, A. & E. Stancanelli (2010): "Does Income Taxation Affect Partners Household Chores ?" *IZA Working Paper* **1(5038)**