Title
Essays in Household Economics

Author
Fon, Alexandre

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Essays in Household Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Alexandre Fon

2021

ABSTRACT OF THE DISSERTATION

Essays in Household Economics

by

Alexandre Fon
Doctor of Philosophy in Economics
University of California, Los Angeles, 2021
Professor Maurizio Mazzocco, Chair

This dissertation contains three essays in applied microeconomics, with a focus on household decision-making.

In the first chapter, I study the effect of asymmetric information about income on household decisions, resource sharing, and welfare. I proceed in four steps. In the first step, I develop a theoretical model that accounts for the possible existence of asymmetric information. The model predicts that households will partly mitigate the welfare cost of asymmetric information by incentivizing the wage earner to provide information about his or her true income. These incentives are provided by making the consumption share increase with reported income: the wage earner's consumption share is high when reporting a high income and low when reporting a low income. Second, I derive a new non-parametric identification result for this model. Third, I estimate the model using a survey of Bangladeshi day laborers. The estimation confirms the predictions of the model, providing evidence that the households in the data are affected by asymmetric information. Finally, I conduct three counterfactual analyses to document how asymmetric information interacts with policies and compute the willingness to pay in each case.

In the second chapter, which is co-authored with Maria Casanova and Maurizio Mazzocco, we show that the intratemporal and intertemporal preferences of each decision-maker in the household can be identified even if individual consumption is

not observed. This identification result is used jointly with the Consumer Expenditure Survey (CEX) to estimate the intratemporal and intertemporal features of individual preferences. The empirical findings indicate that there is heterogeneity in intertemporal preferences between wife and husband.

In the third chapter, I use a major reform of the parental leave system in Quebec in 2006 to analyze how households make decisions related to parental leave. I show that the introduction of a father's quota - a policy designed to incentivize fathers to take parental leave - was successful in more than doubling the proportion of fathers taking some parental leave. However, the impact on the intensive margin was limited: in 80% of households, mothers take all the leave that is available to both parents. I also use an administrative dataset to analyze the relationship between parental leave decisions and income. In general, households with higher labor income take more parental leave overall (summing the mother's and the father's weeks). However, fathers with higher labor income take less parental leave.

The dissertation of Alexandre Fon is approved.

Sarah J. Reber

Bernardo S. Da Silveira

Rosa Liliana Matzkin

Maurizio Mazzocco, Committee Chair

University of California, Los Angeles

2021

À mes parents, pour m'avoir transmis leur curiosité et leur amour de la découverte.
And to Mead for sharing her smile and optimism with me every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Maurizio Mazzocco, for his constant support and invaluable advice, which made this dissertation possible. Our frequent discussions were stimulating and challenging, and greatly shaped me as an economist. I could not have asked for a better advisor during my graduate studies. I am grateful to Bernardo Silveira, Rosa Matzkin, and Sarah Reber for their help and their guidance in completing this dissertation. I am also thankful for the time and advice of other faculty, including Moshe Buchinsky, Daniel Haanwinckel, Adriana Lleras Muney, and Moritz Meyer-ter-Vehn.

Sharing my time at UCLA with exceptional classmates made my experience considerably more enriching, both on an academic and on a human level. I thank Sepehr Ekbatani, Manos Hatzikonstantinou, and Alex Graupner for being such great friends since the first day in Los Angeles. Thanks to Vladimir Pecheu and Rustin Partow, I would look forward to learning something new every time in our weekly reading group. I am also grateful to many other colleagues and friends, including Tomás Guanziroli, Ariadna Jou, Jon Gu, Tina Cui, Carolina Arteaga, Hadi Çaoui, Renato Giroldo, Ksenia Shakhgildyan.

I thank my family - my parents, my sister, my brother, my grand-parents - for their constant encouragement and for always being there for me. Finally, I thank Mead Baldwin for being incredibly supportive and positive.

VITA

Degrees

2015	M.A. in Economics, Sciences Po Paris.
2013	B.A. in Social Sciences, Sciences Po Paris.

Employment

2016-2020	Teaching Assistant, UCLA.
2018	Research Assistant, UCLA.

Honors and Awards

2020-2021	UCLA Graduate Dissertation Year Fellowship.
2020	Bank of Canada Graduate Paper Award Short-List.
2020	UCLA Award Recognizing Outstanding Performance in the Shift to Remote Learning.
2017	UCLA Graduate Summer Research Mentorship.

CHAPTER 1

Asymmetric Information and Hidden Income in Households, with an Application to Bangladesh

Alexandre Fon[1]

1.1 Introduction

Analyzing the effect of policies on household's decisions requires considering the preferences and behavior of individual household members. The collective model (Chiappori 1988, 1992) is an effective and widely used way of doing precisely that. The collective model characterizes the household as a group of individuals, with possibly different preferences, making joint and efficient decisions. It has been used to analyze a large number of important economic questions, ranging from the optimal design of taxation systems and social safety nets, to poverty reduction in developing countries, to the education and labor force participation choices of women. To give a few examples, Low, Meghir, Pistaferri, Voena (2018) use the collective model to analyze PRWORA a major reform to the welfare system in the United States; Gayle and Shephard (2019) study optimal income taxation and its relationship with family structure; Attanasio and Lechene (2014) study PROGRESA, a conditional cash transfer to poor households in Mexico;

[1]Website: https://alexandrefon.io. E-mail address: afon@ucla.edu. I am deeply indebted to my advisor, Maurizio Mazzocco. I thank Bernardo Silveira, Rosa Matzkin, Sarah Reber, Moshe Buchinsky, Adriana Lleras-Muney for their feedback and guidance. I also thank my colleagues Rustin Partow, Vladimir Pecheu, Emmanouil Chatzikonstantinou, Sepehr Ekbatani and seminar participants at the Bank of Canada for insightful comments and discussions. All errors are my own.

Dunbar, Lewbel and Pendakur (2013, 2019) show that standard poverty indices can underestimate child poverty; Voena (2015) and Chiappori, Fortin and Lacroix (2002) show the link between marital contracts and female labor participation; Bronson (2015) uses the collective model to explain the increased college attendance of women since the 1970s. These examples highlight the prevalence and importance of the collective model and, as a consequence, how crucial it is to carefully examine the assumptions underlying the model.

An important assumption underlying the collective model is that household members have complete information about each other's income. However, there is evidence that, for households in developing countries, asymmetric information is a source of inefficiency in the household. Field experiments, such as the one conducted by Ashraf (2009) in the Philippines, document that household members spend cash transfers differently depending on whether or not their spouse knows about that cash transfer. Experiments have found similar results in many developing countries (e.g., Castilla and Walker (2013), Hoel (2014), Ambler (2015), and Boltz, Marazyan, Villar (2019)). This evidence indicates that if the standard collective model is used to answer policy questions, in particular in developing countries, it may lead to misleading conclusions. More generally, the evidence indicates the need for a framework to better understand how asymmetric information affects household's decisions.

The main contribution of this paper is to provide such a framework and to use it to study the effect of asymmetric information about income on household decisions, resource sharing, and welfare. In order to do this, this paper proceeds in four steps. First, it develops a model of a household facing asymmetric information about the income of the wage earner. The purpose of the model is to provide a framework to predict the effect of asymmetric information on household outcomes, test for the presence of asymmetric information, and conduct counterfactual analyses. Second, a new identification result, specific to this model, states that individual welfare functions can be estimated from consumption data. Third, the model is estimated using a sample of Bangladeshi day laborers. The estimation provides evidence that the day laborers are affected by asymmetric information. Finally, I conduct three counterfactual analyses to document how asymmetric information interacts with policies and compute the willingness to pay in each case.

In the model introduced in this paper, households are composed of wage earners with short-term income shocks that are unobservable to other household members. While households make efficient decisions, they are constrained by asymmetric information: the households' equilibrium allocations must be incentive compatible with the wage earners' preferences. Thus, the model relaxes and nests the complete information assumption of the collective household model (Chiappori 1992) in order to account for the evidence on asymmetric information provided by previous papers.

The model shows that households making efficient decisions can reduce the cost of asymmetric information by incentivizing the wage earner to report their true income. This incentive is provided by making the individual consumption share of the wage earner increasing in the reported income. More precisely, if the wage earner reports a low income, they will get a smaller share of household consumption and if they report a high income they will get a bigger share of household consumption. As a consequence, they have an incentive to reveal their income. This reduces the welfare cost of asymmetric information. However, there will still be some cost caused by risk not being shared efficiently because the individual consumption of the wage earner responds "too much" to individual short-term income shocks. The main prediction that emerges from the model is therefore that, under asymmetric information, the consumption share of the wage earner will respond more to short-term income shocks than under complete information.

The main modeling choice underlying this prediction is that households make efficient decisions. The reason for modeling households in this way is that many papers test and fail to reject efficiency in developing countries. Most relevant for the context of my empirical results, Bargain, Lacroix, Tiberti (2018) document that data from Bangladeshi households are consistent with efficient decision-making. Other examples of papers that fail to reject efficiency in developing countries include Attanasio and Lechene (2014) and Bobonis (2009). While these tests are designed with complete information in mind, the tests are still valid under asymmetric information. The reason is that, under asymmetric information, income realizations act as a shift in the decision power, but allocations are ex-post Pareto efficient for every realization. Altruism between household members and coordination through repeated interactions are the main reasons that explain why households behave efficiently.

In order to bring the model to the data, a new nonparametric identification result specific to this model shows how individual welfare functions can be recovered from consumption data. In the data, consumption is typically observed at the household rather than individual level. This is the main challenge to identification in this literature. In this paper, I show that the model is identified if two conditions are satisfied. First, identification requires at least one "assignable good" – a good observed at the individual level. Second, identification requires observing both average income, which is common knowledge in the household, as well as some measure of short-term income variation that can be affected by asymmetric information. The identification result also allows the effect of asymmetric information to be separately identified from other factors impacting household members' consumption shares.

The identification result is then used to estimate the model using a sample of Bangladeshi day laborers and their households. This data set is particularly suitable for the analysis for two reasons. First, families with day laborers are more likely to be affected by asymmetric information. Indeed, day laborers' jobs are often informal and have high short-term income variation, making it difficult for other household members to verify the day laborer's daily income. Second, the dataset contains two crucial variables for estimation. The first is food consumption at the individual level, which will serve as an assignable good. The second is short-term income variation, which is likely not to be observed by the spouse in this setting.

The estimation provides evidence that the households in the data are affected by asymmetric information. Since the model in this paper nests the case with complete information, I can test whether complete information is rejected by the data. The estimation shows that short-term income shocks lead to an increase in the day laborer's individual consumption share. This implies that household members are not sharing risk efficiently and allows me to reject complete information. In order to confirm the asymmetric information mechanism, I document that short-term income shocks that are likely to be less observable by the spouse have a larger effect on individual consumption. Specifically, I find that income shocks of day laborers that work outside their village have a bigger impact. I also test the alternative hypothesis that the variation in consumption share is driven by nutritional needs: day laborers that work more have higher income and higher nutritional needs, which would lead to an increased consumption share. I do this by dividing income variation between daily wage variation and days worked variation. I find that income shocks due to variation in wages have a larger impact than variation in days worked, which provides evidence that the results are not being driven exclusively by nutritional needs. Overall, the evidence supports the model and the specific informational mechanism.

The estimation of the model makes it possible to conduct counterfactuals. The first counterfactual analysis quantifies the extra welfare cost if households do not make efficient decisions and do not incentivize truthful income reporting. The goal of this counterfactual is to help interpret past experimental results on asymmetric information. Experiments are one-off situations, which household members might not have experience navigating. As a result, households in experiments might not be able to provide incentives to mitigate the effect of asymmetric information. Therefore, experimental results might overstate the effect of asymmetric information compared to when it is a repeated issue facing the household. The simulation documents that without incentives, the day laborer chooses to hide 44% of the unobserved income. The consumption share of the wage earner varies considerably more in response to a short-term income shock without incentives to report truthfully: an income shock 10% above average income increases the wage earner's consumption share by 4.4% without incentives and only 0.14% with incentives. Therefore, experiments might be considerably overstating the cost of asymmetric information.

I then conduct two policy counterfactuals. First, I consider a guaranteed employment scheme. This policy is used in developing countries, such as Bangladesh and India, to help poor workers through periods of low employment. The daily wage of participants is constant and fixed nationally, and therefore is not likely to be affected by asymmetric information. A policy simulation reveals that households are willing to pay 0.3% of total yearly household income to remove asymmetric information through a guaranteed employment scheme for a one year period. This indicates that, while asymmetric information does affect households, the welfare cost is quite small. Households are able to limit the welfare effects of asymmetric information by providing incentives to report income truthfully. Second, I consider a tax on a good that can be consumed with hidden income. I show that a 50% tax on such a good reduces the "extra" variance in consumption share due to asymmetric information from 26% to 16% of the total variance in consumption share.

This paper contributes to three main strands of literature. First, this paper contributes to the literature on cooperative models of intra-household allocations. In particular, it extends the standard collective model of the household, which was first introduced by Chiappori (1988, 1992) and which is comprehensively reviewed by Chiappori and Mazzocco (2017), to allow for asymmetric information. In addition, identification results from this literature (Chiappori and Ekeland 2009; Blundell, Chiappori, Meghir 2005) are extended to the model in this paper. More broadly, this paper contributes and builds on the literature on estimating collective models. The empirical specification used to estimate the model is closest to the one in Cherchye, De Rock, Vermeulen's (2012) paper. Other papers that have structurally estimated the collective model include Browning, Bourguignon, Chiappori and Lechene (1994); Dunbar, Lewbel, and Pendakur (2013, 2019); Mazzocco, Ruiz, and Yamaguchi (2014).

Second, this paper contributes to the literature on risk-sharing within the household. This paper provides evidence that households do not share risk efficiently. This is consistent with the results of Dercon and Krishnan (2000) and Duflo and Udry (2004) that find that poor households do not share risk efficiently in Ethiopia and the Ivory Coast respectively. My paper highlights the importance of asymmetric information in explaining this surprising result. In contrast to papers finding imperfect insurance, Haushofer and Shapiro (2014) and Benhassine et al. (2015) find no evidence that the gender of the recipients of a cash transfer affects household outcomes. However, in both cases the cash transfers are given in the context of a large scale field experiment, and therefore are likely to be common knowledge in the household. Dubois and Ligon's (2002) paper is closest to this paper. They also document that consumption shares respond to short-term income shocks and test whether asymmetric information is the cause. They conclude that asymmetric information is partly responsible for the observed variation in consumption shares. This paper differs along several dimensions. First, the source of asymmetric information Dubois and Ligon highlight concerns the effort of household members when working, while this paper focuses on asymmetric information about income. As a result, the testable implications that emerge from the two models are different. Second, this paper not only tests for asymmetric information, but also identifies and estimates the whole model, which then allows counterfactual analyses to be conducted. Mazzocco (2007) proposes a model with imperfect insurance within the household due to limited commitment of household members to future outcomes: household members renegotiate their share of resources if their outside option of leaving the household increases enough. While limited commitment can explain imperfect insurance for larger income shocks, short-term income shocks are unlikely to significantly impact the outside option. Therefore, the asymmetric information mechanism proposed in this paper is likely to be the more relevant mechanism for imperfect insurance of short-term income shocks.

Third, many papers document the impact of asymmetric information about income on household behavior in experimental settings. For instance, Ashraf (2009) shows that individuals will spend windfall income differently depending on whether their spouse knows about the income. In this situation, one of the spouses takes advantage of the asymmetric information to consume goods that are more privately beneficial than if the income had been shared with the other spouse. Other papers also find evidence that asymmetric information matters: Castilla and Walker (2013); Hoel (2014); Ambler (2015). These papers provide very useful evidence that asymmetric information matters for household decisions. My analysis predicts that households that face asymmetric information repeatedly (like those with day laborers) can reduce the cost of asymmetric information through incentives. By contrast, experiments are typically one-offs and situations household members have no experience dealing with. As a result, household members are likely not able to coordinate to provide incentives in an experiment, as they would in a repeated setting outside of the lab. Therefore, the experimental evidence might be overestimating the impact of asymmetric information. An added benefit of the modeling approach I take is that I can estimate the model, which makes it possible to run counterfactuals.

The rest of the chapter is organized as follows: the second part develops the model of the household. The third part derives an identification result for this model. The fourth part describes the data. The fifth part specifies a parametric empirical specification in order to estimate the model. The sixth part estimates the model and presents further results and counterfactuals. Finally, the seventh part concludes.

1.2 Model

1.2.1 Setting

This section describes the economic setting of a cooperative household that faces asymmetric information about income. The household is a two person household acting cooperatively. Agent 1, the wage earner draws random income $Y = \bar{y} + \tilde{y}$ where $\tilde{y}$ comes from a continuous distribution $F(\tilde{y})$ that is common knowledge. The income realization is not observed by agent 2, the home producer. Agent 1 has the possibility to consume a consumption good h_1 before revealing his income. The consumption of h_1 in this case is not observed by the household, allowing agent 1 to hide income and pretend to have had a lower income realization. The income revealed to the household is then used to consume two private goods (c and x) for each household member (c_1, x_1 for agent 1 and c_2, x_2 for agent 2) and a public good (X). The model can easily be generalized to a larger number of private and public goods. However, for the purpose of identification, which will be discussed later, there needs to be at least one good that is observed at the individual level in the data - c_i here. I will refer to c_i, x_i and X as "within-household" consumption. It is assumed that the preferences of agent i in the household can be represented by the following utility functions: $U^i(c_i, x_i, X) + v_i(h_i)$. In this model, agents have different preferences over h_i and the other private goods. This reflects the fact that hiding income limits consumption. In particular, consumption that would be observed by the spouse cannot be consumed with hidden income. Therefore, we can think that in general, consuming the hidden good has a "cost" reflected in the preferences, which comes from the fact that your consumption options are limited.

The presence of asymmetric information constrains the possible consumption allocations that can be reached in equilibrium. The allocations have to be incentive compatible for agent 1 $-$ i.e., he is weakly better off by revealing his true income than by pretending to have a lower income and consuming the difference in incomes as hidden consumption. Subject to the incentive compatibility constraint, the outcomes are assumed to be Pareto efficient. The household's problem is the following[2]:

[2]A solution to the following problem will be Pareto efficient. However, the converse is not necessarily true. The standard proof is based on the separating hyperplane theorem that applies to convex sets. Here the set of possible allocations - that are both feasible and IC - is not necessarily convex because

$$\max_{c_i,x_i,h_i,X,i=1,2\ \forall\tilde{y}} \mathbb{E}_{\tilde{y}}\ \mu(\bar{y},P,\pi)[U^1(c_1,x_1,X)+v_1(h_1)]+(1-\mu(\bar{y},P,\pi))[U^2(c_2,x_2,X)+v_2(h_2)] \tag{P1}$$

$$\text{s.t.} \qquad c_1+c_2+px_1+px_2+\pi X+qh_1+qh_2=\bar{y}+\tilde{y} \qquad \forall\tilde{y}$$

$$U^1(c_1,x_1,X)+v_1(h_1)\geq U^1(c_1',x',X')+v_1(h_1'+\frac{\tilde{y}-\tilde{y}'}{q}) \quad \forall\tilde{y}'<\tilde{y}$$

In the above, the first constraint is the budget constraint and the second constraint is the IC constraint. The IC constraint says that agent 1 must weakly prefer the equilibrium allocation (on the left-hand side) to deviating by pretending to have a lower income draw and spending the difference in income on the outside good (on the right-hand side). p is the price of good x, π is the price of the public good and q the price of the hidden consumption good. μ the decision power of agent 1. It can depend arbitrarily on the average income $\bar{y}$, the vector of private prices $P=(p,q)$ and π. Here it is assumed that agent 1 is egoistic (the IC constraint depends only on the preferences of agent 1). However, it is easy to add "separable" altruism in this model: agent 1's utility is a weighted sum of his and agent 2's egoistic utility functions. The predictions of the model would be unchanged by adding this specific form of altruism.

There is a slight abuse of notation in the problem above. Technically, each decision variable is a function of $\tilde{y}$ since we have to solve the problem for each realization of income. We suppress the function for ease of notation, but it is important to remember that we are solving for a function for each of these decision variables. Note that if it were not for the IC constraint, the household's problems for any two realizations of $\tilde{y}$ would be independent. The IC constraint connects the problem for two realizations and makes it such that maximizing the household's expected utility is more difficult than simply maximizing the utility at every realization of $\tilde{y}$. Such a naive approach would not respect the IC constraint and, therefore, agent 1 would have an incentive to deviate.

---

of the IC constraint.

1.2.2 IC constraint

The IC constraint in this specific form is a complicated object. Therefore, we can manipulate it to make it easier to deal with. Let us denote $IC(\tilde{y}, \tilde{y}')$ the IC constraint for a specific $\tilde{y}$ and $\tilde{y}'$ with $\tilde{y} > \tilde{y}'$. Under the assumption that $v(.)$ is weakly concave, it is possible to show that for any $\tilde{y} > \tilde{y}' > \tilde{y}''$, $IC(\tilde{y}, \tilde{y}')$ and $IC(\tilde{y}', \tilde{y}'')$ imply $IC(\tilde{y}, \tilde{y}'')$. Therefore, if the IC constraint holds locally, it holds globally. In other words, if agent 1 does not have an incentive to pretend to have income slightly below his true income he does not have an an incentive to pretend to have any other income. If the distribution of income is discrete, any realization of income $\tilde{y}$ appears only in two binding IC constraints: with the value just above $\tilde{y}$ and with the value just below $\tilde{y}$. Since here the distribution of income is continuous, the IC constraint will constrain the derivative of agent 1's indirect utility. To show this, let us rewrite the IC constraint above in terms of the utility level of within-household consumption when reporting income $\tilde{y}$, $\tilde{U}^1(\tilde{y}) \equiv U^1(c_1(\tilde{y}), x_1(\tilde{y}), X(\tilde{y}))$:

$$\tilde{U}^1(\tilde{y}) + v_1(h_1(\tilde{y})) \geq \tilde{U}^1(\tilde{y}) + v_1\big(h_1(\tilde{y}') + \frac{\tilde{y} - \tilde{y}'}{q}\big) \qquad \forall \quad \tilde{y} > \tilde{y}'$$

We can divide on both sides by $\tilde{y} - \tilde{y}'$. Rearranging, we get:

$$\frac{\tilde{U}^1(\tilde{y}) - \tilde{U}^1(\tilde{y}')}{\tilde{y} - \tilde{y}'} \geq -\frac{v_1(h_1(\tilde{y})) - v_1(h_1(\tilde{y}') + \frac{\tilde{y}-\tilde{y}'}{q})}{\tilde{y} - \tilde{y}'} \qquad \forall \quad \tilde{y} > \tilde{y}'$$

Taking the limit as $\tilde{y}'$ goes to $\tilde{y}$, we get

$$\tilde{U}'^1(\tilde{y}) \geq v_1'(h_1(\tilde{y}))\big(\frac{1}{q} - \frac{\partial h_1(\tilde{y})}{\partial \tilde{y}}\big)$$

We can rewrite this in an intuitive manner:

$$\tilde{U}'^1(\tilde{y}) + v_1'(h_1(\tilde{y}))\frac{\partial h_1(\tilde{y})}{\partial \tilde{y}} \geq \frac{1}{q} v_1'(h_1(\tilde{y})) \tag{1.1}$$

This equation states that the marginal utility from revealing true income has to be equal to or greater than the marginal utility from hiding income for agent 1. The marginal utility from revealing true income is the sum of the marginal utility from within-household consumption and the marginal utility from the change in hidden consumption in the equilibrium allocation. Meanwhile, the marginal utility from hiding income is simply the increased utility from consuming only the hidden consumption good with the hidden income.

Integrating equation (1.1) on both sides with respect to $\tilde{y}$ between some arbitrary $\tilde{y}$ and $\tilde{y}_0$ (the smallest possible value) we get the following, where $U_0^1 = U^1(c_1(\tilde{y}_0), x_1(\tilde{y}_0), X(\tilde{y}_0)) + v_1(h_1(\tilde{y}_0))$ is agent 1's utility when he truthfully reports the lowest income realization $\tilde{y}_0$:

$$U^1(c_1(\tilde{y}), x_1(\tilde{y}), X(\tilde{y})) + v_1(h_1(\tilde{y})) - U_0^1 \geq \int_{\tilde{y}_0}^{y} v_1'(h_1(t))\mathrm{d}t \tag{1.2}$$

Note that in the special case where $v_1(h_1)$ is linear, the RHS of equation (1.2) becomes a linear function of $\tilde{y} - \tilde{y}_0$. In that case, the IC constraint will restrict agent 1's utility to increase linearly in income. The more general case in equation (1.3) is similar. The main difference is that the slope of agent 1's utility is not constant, but rather decreases with $h_1(\tilde{y})$ (under the assumption that $v_1(.)$ is concave).

For the remainder, I will assume that the IC constraint binds for all $\tilde{y}$. Therefore, I will treat equation (1.3) as an equality.

1.2.3 Solution

We can therefore rewrite the household's problem as:

$$\max_{c_i, x_i, h_i, X,\ \forall\tilde{y}} \ \mathbb{E}_{\tilde{y}} \ \mu(\bar{y}, P, \pi)[U^1(c_1, x_1, X) + v_1(h_1)] + (1 - \mu(\bar{y}, P, \pi))[U^2(c_2, x_2, X) + v_2(h_2)]$$

$$\text{s.t.} \quad c_1 + c_2 + px_1 + px_2 + \pi X + qh_1 + qh_2 = \bar{y} + \tilde{y} \qquad \forall \tilde{y}$$

$$U^1(c_1, x_1, X) + v_1(h_1) \geq U_0^1 + \frac{1}{q}\int_{y^0}^{\tilde{y}} v_1'(h_1(t))\mathrm{d}t$$

We can take FOCs to solve this problem. λ_1 and λ_2 are the Lagrange multipliers associated with the budget constraint and the IC-constraint respectively. I will denote by U_k^i the derivative of the utility function of agent i with respect to argument k. I also will not write the decision variables as functions of y explicitly for ease of notation

$$(\mu + \lambda_2)U_1^1(c_1, x_1, X) = \lambda_1 \tag{c_1}$$

$$(1 - \mu)U_1^2(c_2, x_2, X) = \lambda_1 \tag{c_2}$$

$$(\mu + \lambda_2)U_2^1(c_1, x_1, X) = \lambda_1 \tag{x_1}$$

$$(1 - \mu)U_2^2(c_2, x_2, X) = \lambda_1 \tag{x_2}$$

$$(\mu + \lambda_2)U_3^1(c_1, x_1, X) + (1 - \mu)U_3^2(c_2, x_2, X) = \pi\lambda_1 \tag{X}$$

$$(\mu + \lambda_2) v_1'(h_1(y)) = q\lambda_1 + \lambda_2 \frac{1}{q} v_1'(h_1(y)) \frac{1}{h_1'(y)} \tag{h_1}$$

$$(1-\mu) v_2'(h_2) = q\lambda_1 \tag{h_2}$$

This set of equations defines the allocation for each income realization $\tilde{y}$. The first thing to note is that this set of equations is very similar to the set of equations we would have if there was no IC constraint. In fact, apart from the FOC for h_1, the other equations are exactly the same that we would get if we solved the household's problem with Pareto weight on agent 1 being $\mu + \lambda_2$ and on agent 2 being $1 - \mu$. Normalizing the Pareto weight to sum to 1, this would imply that the Pareto weight of agent 1 will be $\frac{\mu+\lambda_2}{1+\lambda_2}$ instead of μ. Therefore, for all goods apart from h_1 the IC constraints can be thought of as simply shifting the decision power of agent 1. This can be seen most clearly in the modified efficiency rule:

$$\frac{\mu + \lambda_2}{1 - \mu} = \frac{U_1^2(c_2, x_2, X)}{U_1^1(c_1, x_1, X)}$$

Under complete information, we would have $\lambda_2 = 0$ for all income realizations. However, with asymmetric information, the ratio of marginal utilities will deviate from the ratio of Pareto weights.

Looking at equation (1.2), we can see that with a concave $v(.)$, changing consumption of hidden income can make the IC-constraint slacker or tighter. Therefore, it is to be expected that the optimal consumption of h_1 would take this effect into account. More precisely, what this tells us is that the marginal benefit of consuming h_1 has to equal to marginal cost. This marginal cost has the usual BC term $q\lambda_1$ and an additional term. When λ_2 is negative, this cost is also negative, therefore the wage earner will consume more h_1. The inverse happens when λ_2 is positive.

An important object in this set of equations is λ_2. $\lambda_2/(1 + \lambda_2)$ is the additional decision power of agent 1 for any given income realization. Solving for λ_2, we find that:

$$\lambda_2 = \frac{(1-\mu)U_1^2(c_2, x_2, X) - \mu U_1^1(c_1, x_1, X)}{U_1^1(c_1, x_1, X)} = -\mu + (1-\mu)\frac{U_1^2(c_2, x_2, X)}{U_1^1(c_1, x_1, X)} \tag{1.3}$$

λ_2, the multiplier on the IC constraint, has an intuitive form. We have seen that the choice of c_1, x_1 and X is efficient for any given realization of y, with only the decision power shifting. Therefore, λ_2 is simply how much the ratio of marginal utilities deviates from the standard efficiency condition. Note that λ_2 can be positive or negative depending on which marginal utility is "too high."

For a given U_0, the set of FOCs above, along with the budget constraint and the IC constraint, define the optimal allocation. Therefore, to completely characterize the optimal allocation solving for U_0 is the only remaining step.

Taking the FOC for U_0 we get:

$$\mathbb{E}_y[-\lambda_2] = 0$$

Plugging in the expression for λ_2 given in equation (1.3) and rearranging, we get:

$$\mathbb{E}_y\left[\frac{U_1^2(c_2, x_2, X)}{U_1^1 c_1, x_1, X}\right] = \frac{\mu}{1-\mu}$$

Therefore we find that even though the efficiency rule does not hold at each $\tilde{y}$, U_0^1 is chosen such that it holds on average.

1.2.4 Two-Stage Decision Process

A helpful tool to think about identification is to break down the household's decision in a two-stage decision process. It allows us to define a *conditional sharing rule* (this idea is standard in collective household models, for example Blundell, Chiappori, Meghir (2005)). The sharing rule summarizes how household members share resources and, therefore, selects one particular outcome out of the set of constrained efficient outcomes. In that capacity, it serves the same purpose as the Pareto weights but with the added benefits of being more clearly interpretable and not depending on the particular cardinalization of utilities.

The two stages will be the following. In the first stage, the household jointly chooses U_0^1, h_i, X and how to split the residual income between them. In the second stage, each household member freely chooses how to allocate this income between the two private goods c_i and x_i. More formally, denote by $c_i(P, \pi, \bar{y}, \tilde{y})$, $x_i(P, \pi, \bar{y}, \tilde{y})$, $h_i(P, \pi, \bar{y}, \tilde{y})$, $X(P, \pi, \bar{y}, \tilde{y})$ the solution to the household's problem. Then define the sharing rule: $r(P, \pi, \bar{y}, \tilde{y}) = c_1(P, \pi, \bar{y}, \tilde{y}) + px_1(P, \pi, \bar{y}, \tilde{y})$. Then we will have:

$$r^1(P, \pi, \bar{y}, \tilde{y}) \equiv r(P, \pi, \bar{y}, \tilde{y})$$

$$r^2(P, \pi, \bar{y}, \tilde{y}) \equiv \bar{y} + \tilde{y} - \pi X(P, \pi, \bar{y}, \tilde{y}) - q(h_1(P, \pi, \bar{y}, \tilde{y}) + h_2(P, \pi, \bar{y}, \tilde{y})) - r(P, \pi, \bar{y}, \tilde{y})$$

from the budget constraint. The r^i functions simply tell us how much of the residual income after consuming the public good and the hideable good is given to each household member i.

Proposition 1. Let $c_i^* = c_i(P, \pi, \bar{y}, \tilde{y})$, $x_i^* = x_i(P, \pi, \bar{y}, \tilde{y})$, $h_i^* = h_i(P, \pi, \bar{y}, \tilde{y})$, $X^* = X(P, \pi, \bar{y}, \tilde{y})$ be the solution to the household's problem defined in $(P1)$. Then $c_i(P, \pi, \bar{y}, \tilde{y})$, $x_i(P, \pi, \bar{y}, \tilde{y})$ solves:

$$\begin{aligned} \max_{c_i, x_i} \quad & U_i(c_i, x_i, X^*) \\ \text{s.t.} \quad & c_i + px_i = r^i(P, \pi, \bar{y}, \tilde{y}) \end{aligned} \tag{1.4}$$

Proof for agent 2. For agent 2 this is a very simple proof by contradiction. Suppose that the proposition does not hold for agent 2. Then there must be a c_2', x_2' that respects the budget constraint such that $U_2(c_2', x_2', X^*) > U_2(c_2^*, x_2^*, X^*)$. However, if that were the case then the maximand in problem $(P1)$ could be increased by replacing c_2^* and x_2^* by c_2' and x_2' while still satsifying all the constraints, which is a contradiction. □

For agent 1, we just need to be a little more careful because the utility for a given $\tilde{y}$

is constrained by the IC constraint. In particular, an increase in the utility of agent 1 at a given income realization might not be feasible in the household's problem because of the IC constraint. However, the separability[3] of outside consumption h_1 means that only the within-household utility level matters (U_1^*) for the IC constraint and not which goods are being consumed. Therefore, agent 1 will be optimizing in the second stage.

Proof for agent 1. Suppose that the proposition does not hold for agent 1. Then there must be a c_1', x_1' that respects the budget constraint such that $U_1(c_1', x_1', X^*) > U_1(c_1^*, x_1^*, X^*)$. Then, since utilities are increasing in the various arguments, there must a c_1'', x_1'' such that $U_1(c_1'', x_1'', X^*) = U_i(c_1^*, x_1^*, X^*)$ but $c_1'' + px_1'' < r_1^*$. Then, we have leftover income that we can give to agent 2. For example, the bundle (c_1'', x_1'', $c_2'' = c_2^* + (r_1^* - (c_1'' + px_1''))$, x_2^*, X^*, h_1^*, h_2^*) will increase the maximand from the household's problem. In addition, the budget constraint holds by construction since the extra consumption for agent 2 corresponds precisely to the left over income from agent 1. Finally, the IC constraint will hold since it depends only on c_1 and x_1 through the within-household utility level of agent 1 (U_1^*) which has not changed. Therefore, the maximand of the household's problem can be increased while still respecting all constraints, which is a contradiction. □

1.2.5 Indirect Utilities

We are going to define two relevant indirect utility concepts here. Again, this follows Blundell, Chiappori and Meghir (2005) and Chiappori and Ekeland (2009). The first indirect utility is called the conditional individual indirect utility. It is the maximized value of program number (1.4) for some arbitrary values of r^i and X. Therefore, it is the value of agent i's maximized utility as a function of p, r^i and X and it is denoted $V^i(p, r^i, X)$. It is called "individual" because it depends only on agent i's preferences. It is called "conditional" because it depends on X.

Obviously, if we know individual preferences $V^i(p, r^i, X)$ and the decision process $r^i(P, \pi, \bar{y}, \tilde{y})$ we know everything there is to know about the household. However, as we

[3]Here we have strong separability, but weak separability would be enough for the identification proof to hold.

shall see, it is not possible to identify these two functions separately.

Instead of conditional individual indirect utilities, the identification will aim to recover the conditional collective indirect utilities. This indirect utility is called "collective" because it combines individual preferences and the decision process. It is still called conditional because it still depends on X.

To get this function, two changes of variables are required. The first change of variable is specific to the setting with asymmetric information. One specificity of this setting is that the outside good h_1 has to be chosen in the first stage because it directly affects the IC constraint and not only through agent 1's within-household utility function[4]. Then, notice that h_i enters the definition of the function ρ^2, which will complicate the identification. Therefore, we can make the following change of variables. Denote $y = \bar{y} + \tilde{y} - \pi X(P, \pi, \bar{y}, \tilde{y}) - q(h_1(P, \pi, \bar{y}, \tilde{y}) + h_2(P, \pi, \bar{y}, \tilde{y}))$ the residual income in stage 1 after consuming the public good and the outside good. Then, as long as the partial of residual income with respect to $\bar{y}$ is not 0 [5] we can locally apply the implicit function theorem to express $\bar{y}$ as some function ϕ of the other exogenous variables and residual income. We can then plug ϕ into the sharing rule. We then define the new sharing rule function $\tilde{r}(P, \pi, y, \tilde{y}) = r(P, \pi, \phi(P, \pi, y, \tilde{y}), \tilde{y})$ and we then have $\tilde{r}^1(P, \pi, y, \tilde{y}) = \tilde{r}(P, \pi, y, \tilde{y})$ and $\tilde{r}^2(P, \pi, y, \tilde{y}) = y - \tilde{r}(P, \pi, y, \tilde{y})$. Since residual income is observed directly, this will not complicate identification.

The second change of variable is the standard one in collective models with public consumption. We will vary the price of the public good π such that X is kept constant. Take a neighbourhood of some point $(P, \pi, y, \tilde{y})$ where $\dfrac{\partial X(P, \pi, y, \tilde{y})}{\partial \pi} \neq 0$. By the implicit function theorem we can use the equation $X(P, \pi, y, \tilde{y}) = X$ to get π as some function ψ of all the other exogenous variables and X. Then we can plug this function into the sharing rule $\tilde{r}$ to get the sharing rule as a function of public consumption X.

[4]h_1 has to be chosen in the first stage because the choice of h_1 can make the IC constraint tighter or slacker, affecting the resource split. Therefore, it is similar to a public good because both household members care about the consumption of that good.

[5]It would be 0 only if all the extra income would be used to purchase h_1, h_2 and X, which seems unlikely

Then we can define the conditional sharing rule:

$$\rho(P, y, \tilde{y}, X) = \tilde{r}(P, \psi(P, y, \tilde{y}, X), y, \tilde{y}) \tag{1.5}$$

and the conditional collective indirect utility:

$$W^i(P, y, \tilde{y}, X) = V^i(p, \rho^i(P, y, \tilde{y}, X), X) \tag{1.6}$$

Three things are important to note here. First, we will have that $\rho^1 \equiv \rho(P, y, \tilde{y}, X)$ and $\rho^2(P, y, \tilde{y}, X) \equiv y - \rho(P, y, \tilde{y}, X)$. Second, the collective indirect utility function is the one that is relevant for welfare analysis because it summarizes the utility outcome for agent i of any policy, taking into account the redistribution within the household. Third, since $X(P, \pi, y, \tilde{y})$ and $\psi(P, y, \tilde{y}, X)$ are known functions, it is straightforward to do the change of variable described above in either direction. In particular, if we identify the conditional collective indirect utility it is easy to get the unconditional one by replacing X by the function $X(P, \pi, y, \tilde{y})$.

1.2.6 First Stage of the Household Problem

Now that we haveve defined the indirect utilities from the second stage of the household decision process, we have the tools to write explicitly the first stage. The individual indirect utility is the relevant concept when it comes to writing the first-stage in the two-stage representation of the household's problem. The household decisions in the first stage are made knowing that in the second stage each household member will maximize their utility given the choice of r, X and the value of p. Therefore, the household knows that the choice of X and r will lead to individual utilities $V^i(p, r^i, X)$.

This allows us to write the first stage of the household problem:

$$\begin{aligned}
\max_{r^i, h_i, X, \forall \tilde{y}} \quad & \mathbb{E}_{\tilde{y}} \; \mu(\bar{y}, P, \pi)[V^1(p, r^1, X) + v_1(h_1)] + (1 - \mu(\bar{y}, P, \pi))[V^2(p, r^2, X) + v_2(h_2)] \\
\text{s.t.} \quad & r^1 + r^2 + \pi X + q h_1 + q h_2 = \bar{y} + \tilde{y} \qquad \forall \tilde{y} \\
& V^1(p, r^1, X) + v_1(h_1) \geq V_0^1 + \frac{1}{q} \int_{y^0}^{\tilde{y}} v_1'(h_1(t)) \mathrm{d}t
\end{aligned} \tag{1.7}$$

1.3 Identification

As is common in collective models, a main challenge of identification is that the individual consumption of private goods is not observed separately for all the goods. In particular, in this model, we will assume that we can observe the individual consumption of one of the private goods (c_i with loss of generality) but not of the other (x_i). Rather, for consumption good x we observe only the household aggregate demand $x = x_1 + x_2$. Following the two changes of variables in the previous section, the situation is now very similar to Chiappori and Ekeland (2009). In particular, $\tilde{y}$ can play the role of a distribution factor. Therefore, one assignable good is sufficient to identify the collective indirect utilities.[6] More precisely, the knowledge of c_1, c_2, $x = x_1 + x_2$, X as functions of p, π, y and $\tilde{y}$ is sufficient to recover the welfare-relevant structure. Details are given below.

1.3.1 Identification of the Sharing Rule

Denote $A = \dfrac{\partial c_1(P, y, \tilde{y}, X)}{\partial y} \Big/ \dfrac{\partial c_1(P, y, \tilde{y}, X)}{\partial \tilde{y}}$, $B = \dfrac{\partial c_2(P, y, \tilde{y}, X)}{\partial y} \Big/ \dfrac{\partial c_2(P, y, \tilde{y}, X)}{\partial \tilde{y}}$,

$C = \dfrac{\partial c_1(P, y, \tilde{y}, X)}{\partial y} \Big/ \dfrac{\partial c_1(P, y, \tilde{y}, X)}{\partial q}$ and $D = \dfrac{\partial c_2(P, y, \tilde{y}, X)}{\partial y} \Big/ \dfrac{\partial c_2(P, y, \tilde{y}, X)}{\partial q}$.

[6] In the terminology of the collective model, an assignable good is a good where the consumption of each member is observed separately.

Proposition 2. If $A \neq B$ and $C \neq D$, then the knowledge of the continuous, differentiable c_1, c_2, $x = x_1 + x_2$, X as functions of P, π, y and $\tilde{y}$ allows us to identify the sharing rule up to an additive function of p and X. That is, two sharing rules $\rho(P, y, \tilde{y}, X)$ and $\hat{\rho}(P, y, \tilde{y}, X)$ will be such that $\hat{\rho}(P, y, \tilde{y}, X) = \rho(P, y, \tilde{y}, X) + f(p, X)$

Proof. From the two-stage formulation we have that $c_i(P, y, \tilde{y}, X) = c_i^*(p, \rho^i(P, y, \tilde{y}, X), X)$.

Recall that $\rho^1(P, y, \tilde{y}, Q) = \rho(P, y, \tilde{y}, X)$ and $\rho^2(P, y, \tilde{y}, X) = y - \rho(P, y, \tilde{y}, X)$. Then, from the two-stage formulation, we have that $A = \dfrac{\rho_y}{\rho_{\tilde{y}}}$ and $B = -\dfrac{1-\rho_y}{\rho_{\tilde{y}}}$ where ρ_y denotes the partial of the function $\rho = \rho^1$ with respect to y and similarly for $\tilde{y}$. A and B are known from the knowledge of the demand functions. Solving, we find $\rho_y = \frac{A}{A-B}$ and $\rho_{\tilde{y}} = \frac{1}{A-B}$. Similarly, from C and D we find that $\rho_y = \frac{C}{C-D}$ and $\rho_q = \frac{1}{C-D}$. Note that we are overidentified here since we find two independent expressions for ρ_y[7]. We recovered three of the partial derivatives of ρ. Therefore, if ρ and ρ' are two sharing rules, the difference will be in the form of $f(p, X)$. Alternatively, we have identified ρ up to an additive function of p and X.

□

1.3.2 Identification of the Collective Indirect Utilities

Proposition 3. Under the same conditions as proposition 2, the conditional collective indirect utilities are identified up to an increasing transformation.

Proof. Corollary 8 and therefore proposition 7 from Chiappori and Ekeland (2009) apply here. Proposition 7 gives conditions under which the collective indirect utilities are identified up to an increasing transformation. These conditions hold here.

The first thing to note is that we have exactly the same structure as that paper regarding the within-household utilities W^i. The difference between the two settings

[7] q and $\tilde{y}$ enter this problem similarly to distribution factors in the collective model. This overidentification leads to a natural test of the model that is similar to the well known test of proportionality of distribution factors.

is that this paper has asymmetric information. However, we have already seen that in equilibrium this will affect the allocations of within-household consumptions the same way a variation in the decision power would. In Chiappori and Ekeland, the decision power can vary arbitrarily with the exogenous variables. Therefore, the decision power varying in our setting is not an issue. In addition, thanks to our change of variables the choice of h_i does not affect the second stage. Finally, the exogenous variables that do not have an equivalent in their paper (q and $\tilde{y}$) enter the the second stage only through the sharing rule. Therefore, they are isomorphic to *distribution factors* in their setting.

Since the two settings are equivalent regarding within-household consumption, the same conditions are necessary to apply corollary 8. More precisely, applying that result requires the collective indirect utilities W^i to be "separable through ρ^i". This simply means that we need:

$$\frac{\partial W^i(P,y,\tilde{y},X)}{\partial y}\frac{\partial \rho^i}{\partial \tilde{y}} = \frac{\partial W^i(P,y,\tilde{y},X)}{\partial \tilde{y}}\frac{\partial \rho^i}{\partial y}$$

To see why this holds in this setting, take derivatives of equation (1.6) with respect to y and $\tilde{y}$. Intuitively, this separability simply means that the variables y and $\tilde{y}$ enter each agent's indirect utility through the same function – the sharing rule. Since W^i is separable through ρ^i, we can apply corollary 8 and as a consequence proposition 7 of Chiappori and Ekeland. Therefore the collective indirect utilities are identified. □

This next part will discuss more informally the identification idea. Many of the arguments are adpated from Ekeland and Chiappori (2009) and Bluncdell, Chiappori, and Meghir (2005).

Fix (p, X) for this part of the analysis. Recall that the vector P is the vector of the prices for the private goods (p, q). We have the following:

$$\frac{\partial W^i(P,y,\tilde{y},X)}{\partial \tilde{y}}\Big/\frac{\partial W^i(P,y,\tilde{y},X)}{\partial y} = \frac{\rho^i_{\tilde{y}}}{\rho^i_y} \tag{1.8}$$

$$\frac{\partial W^i(P,y,\tilde{y},X)}{\partial q}\Big/\frac{\partial W^i(P,y,\tilde{y},X)}{\partial y} = \frac{\partial \rho^i}{\partial q}\Big/\frac{\partial \rho^i}{\partial y} \tag{1.9}$$

The right-hand side of the equalities above are known. Therefore, we have identified the ratios of partial derivatives of the collective indirect utility. This is equivalent to identifying the function $W^i(.)$ up to an increasing transformation at a fixed (p, X). Given that this function describes preferences, it is to be expected that we would be able to identify it only up to an increasing transformation. See the appendix for a proof that knowing the ratios of partial derivatives is equivalent to knowing the function up to an increasing transformation.

Note that the collective indirect utility can still vary arbitrarily with p and X. Therefore, the knowledge of $\rho_{\tilde{y}}, \rho_y, \rho_q$ has allowed us to identify the collective indirect utilities W^i up to an increasing transformation that can depend on p and X. Explicitly, at this stage, we know that the relationship between two collective indirect utilities $\tilde{W}^i$ and W^i that respect equations (1.8) and (1.9) must be the following, for some function F^i increasing in its first argument:

$$W^i(P, y, \tilde{y}, X) = F^i(\tilde{W}^i(P, y, \tilde{y}, X), p, X) \tag{1.10}$$

At this stage, we do not know anything about $\frac{\partial W^i(P,y,\tilde{y},X)}{\partial p}$ and $\frac{\partial W^i(P,y,\tilde{y},X)}{\partial X}$. These two partial derivatives are not constrained by equations (1.8) and (1.9). The goal of this next part will be precisely to see what we can learn about these two partial derivatives. Towards this goal let us pick one arbitrary function $\tilde{W}^i$ that respects equations (1.8) and (1.9). Obviously, this will not be an actual collective indirect utility because at this stage we do not know anything about how W^i varies with p and X. The question then becomes: for a given $\tilde{W}^i$, can we recover a unique function F_i (up to an increasing transformation) such that the true collective indirect utility W^i is given by $W^i(P, y, \tilde{y}, X) = F^i(\tilde{W}^i(P, y, \tilde{y}, X), p, X)$? The rest of this section answers positively and describes the steps.

Let us first focus on $\dfrac{\partial W^i(P, y, \tilde{y}, X)}{\partial p}$. By the envelope theorem applied to program (1.4) we find:

$$\frac{\partial V^i(p, r^i, X)}{\partial p} \Big/ \frac{\partial V^i(p, r^i, X)}{\partial r^i} = x_i$$

Now, using the relationship between W^i and V^i described in equation (1.6) we find:

$$\frac{\partial W^i(P,y,\tilde{y},X)}{\partial p}\bigg/\frac{\partial W^i(P,y,\tilde{y},X)}{\partial \tilde{y}} = \frac{x_i+\rho^i_p}{\rho^i_{\tilde{y}}}$$

Then, since $\rho^1 = \rho$ and $\rho^2 = y - \rho$ we have: $\rho_p = \rho^1_p = -\rho^2_p$ and $\rho_{\tilde{y}} = \rho^1_{\tilde{y}} = -\rho^2_{\tilde{y}}$. Therefore, we get the following, where the right hand side is observable and known:

$$\frac{\partial W^1(P,y,\tilde{y},X)}{\partial p}\bigg/\frac{\partial W^1(P,y,\tilde{y},X)}{\partial \tilde{y}} - \frac{\partial W^2(P,y,\tilde{y},X)}{\partial p}\bigg/\frac{\partial W^2(P,y,\tilde{y},X)}{\partial \tilde{y}} = \frac{x_1+x_2}{\rho_{\tilde{y}}} = \frac{x(P,y,\tilde{y},X)}{\rho_{\tilde{y}}} \tag{1.11}$$

Now we want to identify the functions F^i. From equation (1.10) we can express the partial derivatives of W^i as functions of F^i and $\tilde{W}^i$ and replace them into (1.11). Recall that $\tilde{W}_i$ are functions that we have chosen and are therefore known. We get the following:

$$\Big(\frac{\partial F^1}{\partial \tilde{W}^1}\frac{\partial \tilde{W}^1}{\partial p} + \frac{\partial F^1}{\partial p}\Big)\Big/\frac{\partial F^1}{\partial \tilde{W}^1}\frac{\partial \tilde{W}^1}{\partial \tilde{y}} - \big(\frac{\partial F^2}{\partial \tilde{W}^2}\frac{\partial \tilde{W}^2}{\partial p} + \frac{\partial F^2}{\partial p}\Big)\Big/\frac{\partial F^2}{\partial \tilde{W}^2}\frac{\partial \tilde{W}_2}{\partial \tilde{y}} = \frac{x(P,y,\tilde{y},X)}{\rho_{\tilde{y}}}$$

Rearranging, we get the following, where the right hand side is known:

$$\frac{\partial F^1}{\partial p}\bigg/\frac{\partial F^1}{\partial \tilde{W}^1}\Big(\frac{\partial \tilde{W}^1}{\partial \tilde{y}}\Big)^{-1} - \frac{\partial F^2}{\partial p}\bigg/\frac{\partial F^2}{\partial \tilde{W}^2}\Big(\frac{\partial \tilde{W}^2}{\partial \tilde{y}}\Big)^{-1} = \frac{x}{\rho_{\tilde{y}}} - \frac{\partial \tilde{W}^1}{\partial p}\bigg/\frac{\partial \tilde{W}^1}{\partial \tilde{y}} + \frac{\partial \tilde{W}^1}{\partial p}\bigg/\frac{\partial \tilde{W}^2}{\partial \tilde{y}} \tag{1.12}$$

Since utilities can be identified only up to an increasing transformation, the best we can hope for is to identify the ratios $\gamma^i(\tilde{W}^i,p,X) \equiv \dfrac{\partial F^i}{\partial p}\Big/\dfrac{\partial F^i}{\partial W^i}$. Then we must show that the solution to (1.12) in terms of the ratios is unique. Intuitively, since γ^i depends on only three variables but $\dfrac{\partial W^i}{\partial \tilde{y}}$ depends on $(P,y,\tilde{y},X)$ we can expect there to be only one solution. In practice, this will be the case almost always. This is usually described as "generic" identification in the collective model literature. See Blundell, Chiappori and Meghir (2005) for a more precise discussion of genericity.

Let us provide some intuition as to why the solution to (1.12) is unique. Suppose, by way of contradiction, that there are two solutions (γ^1,γ^2) and $(\gamma^{1'},\gamma^{2'})$. Denote

$\delta^i = \gamma^i - \gamma^{i'}$. Then we can take the difference of (1.12) for the two solutions to get:

$$\delta^1(\tilde{W}^1, p, X)\Big(\frac{\partial \tilde{W}^1}{\partial \tilde{y}}\Big)^{-1} - \delta^2(\tilde{W}^2, p, X)\Big(\frac{\partial \tilde{W}^2}{\partial \tilde{y}}\Big)^{-1} = 0 \tag{1.13}$$

Then, for any point at which $\delta^i(\tilde{W}^i, p, X) \neq 0$ then $\delta^j(\tilde{W}^i, p, X) \neq 0$ for $i \neq j$ and we can write (1.13) as:

$$\log \delta^1(\tilde{W}^1, p, X) - \log \delta^2(\tilde{W}^2, p, X) = \log\Big(\frac{\partial \tilde{W}^1}{\partial \tilde{y}}\Big/\frac{\partial \tilde{W}^2}{\partial \tilde{y}}\Big) \tag{1.14}$$

This implies that the right-hand side must be equal to the sum of a function of $(\tilde{W}^1, p, X)$ and a function of $(\tilde{W}^1, p, X)$. This will almost never be the case because in general the partials of $\tilde{W}^i$ depend on all the variables. Therefore, $\delta^1(\tilde{W}^1, p, X)$ and $\delta^2(\tilde{W}^1, p, X)$ must be zero almost everywhere and the solution to (1.12) in terms of γ^i functions is unique. A more precise statement of why the property implied by (1.14) is almost never satisfied can be found in the appendix of Blundell, Chiappori, Meghir (2005).

We have shown that we can generically identify the ratios $\frac{\partial F^i}{\partial p}\Big/\frac{\partial F^i}{\partial W^i}$. If we can also identify the ratios $\frac{\partial F^i}{\partial X}\Big/\frac{\partial F^i}{\partial W^i}$, we will have identified the functions F^i up to an increasing transformation. In fact, recovering this second pair of ratios of partial derivatives will be very similar to the first pair.

From the first stage of the household's problem (1.7) we find that:

$$\frac{\partial V^1(p, r^1, X)}{\partial X}\Big/\frac{\partial V^1(p, r^1, X)}{\partial r^1} + \frac{\partial V^2(p, r^2, X)}{\partial X}\Big/\frac{\partial V^1(p, r^2, X)}{\partial r^2} = \pi$$

Now, using the relationship between W^i and V^i described in equation (1.6) we find:

$$\frac{\partial W^i(P, y, \tilde{y}, X)}{\partial X}\Big/\frac{\partial W^i(P, y, \tilde{y}, X)}{\partial \tilde{y}} = \Big(\frac{\partial V^i(p, r^i, X)}{\partial X})\Big/\Big(\frac{\partial V^i(p, r^i, X)}{\partial r^i}\rho^i_{\tilde{y}}\Big) + \frac{\rho^i_X}{\rho^i_{\tilde{y}}}$$

Then, since $\rho^1 = \rho$ and $\rho^2 = y - \rho$ we have: $\rho_X = \rho^1_X = -\rho^2_X$ and $\rho_{\tilde{y}} = \rho^1_{\tilde{y}} = -\rho^2_{\tilde{y}}$. Also, recall that earlier we defined the function ψ that gives us the price of the public good π as a function of the exogenous variables and the quantity of the public good X.

Therefore, we get the following, where the right-hand side is observable and known:

$$\frac{\partial W_1(P,y,\tilde{y},X)}{\partial X}\Big/\frac{\partial W_1(P,y,\tilde{y},X)}{\partial \tilde{y}} - \frac{\partial W_2(P,y,\tilde{y},X)}{\partial X}\Big/\frac{\partial W_2(P,y,\tilde{y},X)}{\partial \tilde{y}} = \frac{\psi(P,y,\tilde{y},X)}{\rho_{\tilde{y}}} \tag{1.15}$$

Then, use equation (1.10) to express the partials of W_i as a function of F^i and $\tilde{W}^i$, plug the expression into (1.15) and rearrange (this is all exactly the same as for $\frac{\partial F^i}{\partial p}/\frac{\partial F^i}{\partial W^i}$) to get:

$$\frac{\partial F^1}{\partial X}\Big/\frac{\partial F^1}{\partial \tilde{W}^1}\Big(\frac{\partial \tilde{W}^1}{\partial \tilde{y}}\Big)^{-1} - \frac{\partial F^2}{\partial X}\Big/\frac{\partial F^2}{\partial \tilde{W}^2}\Big(\frac{\partial \tilde{W}^2}{\partial \tilde{y}}\Big)^{-1} = \frac{\psi(P,y,\tilde{y},X)}{\rho_{\tilde{y}}} - \frac{\partial \tilde{W}^1}{\partial p}\Big/\frac{\partial \tilde{W}^1}{\partial \tilde{y}} + \frac{\partial \tilde{W}^1}{\partial p}\Big/\frac{\partial \tilde{W}^2}{\partial \tilde{y}} \tag{1.16}$$

Then, this equation will have a unique solution in terms of the ratios $\frac{\partial F^i}{\partial X}/\frac{\partial F^i}{\partial W^i}$. The argument is exactly the same as above. Therefore, we have generically identified the ratios of partial derivatives of the functions F^i. This implies (see appendix) that we have identified the function F^i up to an increasing transformation. Therefore, we have recovered the collective indirect utility up to an increasing transformation.

Identification of Preferences for the Outside Good

The last remaining object is the preferences for the outside good, $v_i(.)$. Suppose for now that we observe both $h_1(P,y,\tilde{y},X)$ and $h_2(P,y,\tilde{y},X)$ separately. Then, we can recover the indirect utilities $v_i^*(P,y,\tilde{y},X) = v_i(h_i(P,y,\tilde{y},X))$ exactly. More precisely, each choice of $W^i(P,y,\tilde{y},X)$ will determine uniquely the function $v_i^*(P,y,\tilde{y},X)$. Since $W^i(P,y,\tilde{y},X)$ is identified up to an increasing transformation, $v_i^*(P,y,\tilde{y},X)$ is also identified up to the choice of that increasing transformation.

While h_2 enters the household's problem only through the preferences of agent 2, h_1 enters through the preferences of agent 1 *and* the IC-constraint. Therefore, identifying preferences for the outside good will be different for each agent. However, for both agents identification will rely on the tradeoff between utility from outside consumption $v_i(h_i)$ and within-household consumption $V^i(p,r^i,X)$. While $V^i(p,r^i,X)$ is not identified, note that from (1.4) we have the following, where $V_2^i(p,r^i,X)$ refers to the partial of $V^i(p, r^i, X)$ with respect to the second argument :

$$V_2^i(p, \rho^i(P, y, \tilde{y}, X), X) = \frac{\partial W^i(P, y, \tilde{y}, X)}{\partial \tilde{y}} \Big/ \rho^i_{\tilde{y}}$$

In the above equation $\rho^i_{\tilde{y}}$ is known. Therefore, each cardinalization of the collective indirect utility $W^i(.)$ will correspond to exactly one $V_2^i(p, \rho^i(P, y, \tilde{y}, X), X)$ function.

The next step is to understand each agent's tradeoff between within-household and outside utility. First, agent 2 will simply equalize the marginal utilities of consumption within the household and outside of the household. The standard condition equating the marginal rate of substitution to the ratio of prices will hold. To see this, simply take the ratio of FOCs with respect to h_2 and with respect to r^2 in the first stage of the household problem defined in (1.5). We get:

$$\frac{v_2'(h_2)}{V_2^2(p, r^2, X)} = q$$

This condition must hold at the solution of the household's problem. Therefore, we can replace h_2 and r^2 by $h_2(P, y, \tilde{y}, X)$ and $\rho^2(P, y, \tilde{y}, X)$ respectively:

$$v_2'(h_2(P, y, \tilde{y}, X)) = qV_2^2(p, \rho^2(P, y, \tilde{y}, X), X) \tag{1.17}$$

Given a cardinalization of $W^i(P, y, \tilde{y}, X)$, the right-hand side of the equation above is known. Therefore, we have identified the left-hand side as well. Finding the partial derivatives of $v_2^*(P, y, \tilde{y}, X)$ is now straightforward. Simply multiply the left-hand side of (1.17) by the relevant partial of $h_2(P, y, \tilde{y}, X)$. That partial is known since we observe $h_2(P, y, \tilde{y}, X)$. For example:

$$\frac{\partial v_2^*(P, y, \tilde{y}, X)}{\partial y} = v_2'(h_2(P, y, \tilde{y}, X)) \frac{\partial h_2(P, y, \tilde{y}, X)}{\partial y}$$

We can recover all the partial derivatives of $v_2^*(P, y, \tilde{y}, X)$ in this way and then simply integrate. Therefore, for a given cardinalization of $W^2(P, y, \tilde{y}, X)$, $v_2^*(P, y, \tilde{y}, X)$ is unique (the integration constant does not affect the household's maximization problem).

For agent 1, the standard condition equating the marginal rate of substitution of within-household consumption and outside consumption to the ratio of prices will not hold. Instead, for agent 1, the tradeoff between within-household consumption and outside consumption will be determined by the IC-constraint. Intuitively, the IC-constraint tells us that the solution must be such that the increase in agent 1's utility from a marginal increase in $\tilde{y}$ equals the increase in utility if agent 1 were to spend that marginal $\tilde{y}$ only on h_1. If that were not the case, agent 1 would be incentivized to hide part of the income shock and spend it on h_1. More formally, consider the first stage of the household problem once again. The IC-constraint must hold at the solution. That is:

$$W^1(P, y, \tilde{y}, X) + v_1(h_1(P, y, \tilde{y}, X)) = V_0^1 + \frac{1}{q}\int_{y^0}^{\tilde{y}} v_1'((P, y, t, X))\mathrm{d}t$$

Take the partial derivative of the IC-constraint above with respect to $\tilde{y}$ to get:

$$\frac{\partial W^1(P, y, \tilde{y}, X)}{\partial \tilde{y}} + v_1'(h_1(P, y, \tilde{y}, X))(\frac{\partial h_1(P, y, \tilde{y}, X)}{\partial \tilde{y}} - \frac{1}{q}) = 0$$

For a given cardinalization of $W^1(P, y, \tilde{y}, X)$, the only unknown in the equation above is $v_1'(h_1(P, y, \tilde{y}, X))$, which we can therefore solve for. Then, just as for agent 2, we can multiply by the relevant partial derivatives of $h_1(P, y, \tilde{y}, X)$ to find all the partial derivatives of $v_1^*(P, y, \tilde{y}, X) = v_1(h_1(P, y, \tilde{y}, X))$. Finally, we can integrate to find the unique $v_1^*(P, y, \tilde{y}, X)$.

1.4 Data

1.4.1 Data Overview

The data used to estimate the model is the Bangladesh Integrated Household Survey. This survey is a large survey of households in Bangladesh. The estimation focuses on households with day laborers because these workers face a great deal of short-term income variation. Therefore, they are particularly likely to be affected by asymmetric information. The sample is restricted to households composed of two adults, the head of the household and the spouse of the head, and their children, grandchildren or nieces and nephews under 25 years old [8]. Only households with at least one child are kept. The purpose of this restriction is to have a homogeneous sample with two main decision makers, which corresponds to the theoretical framework. The restricted sample consists of 1,999 households observed in either 2012 or 2015. In the cases where many household members work as day laborers, the analysis will focus on only one wage earner. Whenever the head of the household works as a day laborer, the analysis will focus on him. In the 4% of cases where the spouse works as a day laborer and the head does not, the analysis focuses on her.

Table 1.1: Day Laborer Summary Statistics

	Mean	SD
Male	0.96	0.20
Age	41.5	10.6
Number of Children	2.25	1.14
No Education	0.64	0.48
Rural	0.94	0.24
Weekly Income ($ US)	12.9	6.81
Weekly Household Income($ US)	18.3	11.6
Days Worked in the Past Week	4.68	1.86
Weekly Hours	36.2	16.8
Works in Agriculture	0.72	0.45
Works In Own Village	0.89	0.31
Observations	1999	

Notes: Summary statistics from BIHS (mean and standard deviation) describing the day laborers and their households. The demographic variables (Male, Age, Number of Children, No Education) refer to the day laborer. Rural is a dummy variable equal to 1 if the household is rural. Weekly Income, Weekly Hours, Works in Agriculture, Days Worked and Works in Own Village refers to the day laborer's job. Weekly Household Income is average total household weekly income.

Table 1.1 presents some summary statistics for the day laborers and the households that compose the main estimation sample. The day laborers are overwhelmingly male. They are part of very poor rural households: the average weekly wage is less than 13 $US, and more than half of them have no education whatsoever. A large majority work as agricultural day laborers. Other typical industries include construction and transport. The typical arrangement in these households (Khandker and Mahmud 2012) is that the head of the household works as a day laborer, while the spouse is in charge of

[8] All household members that are not the head or the spouse of the head are referred to as children from now on.

generating some additional income through home agriculture or raising poultry. While the male wage earner that works outside of the household earns more on average, his income tends to be variable and depend heavily on the seasonal availability of work.

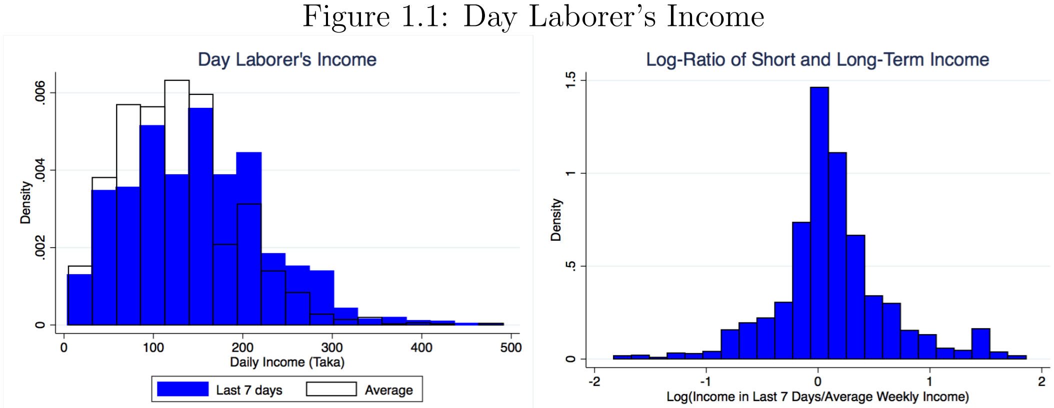
1.4.2 Crucial Data Features and Variable Description

Two non-standard features of the BIHS are going to be crucial to estimate the model. First, for day laborers, the survey reports both average weekly income and income in the seven days preceding the survey. Having these two measures of income is necessary to construct a measure of short-term income shocks, which is the income variation that is potentially affected by asymmetric information. Second, food consumption is reported at the individual level. Therefore, it will serve as the assignable good that is necessary to identify the model. Importantly, individual food consumption is reported in the 24 hours preceding the survey. The model describes how short-term income shocks affect short-term consumption shares in the presence of asymmetric information. Therefore, it is important to use a short-term, non-durable measure of consumption – such as food consumption in the past 24 hours – as the assignable good. Clothing, which has been commonly used as an assignable good in the literature, would not be adequate here.

Three categories of variables are going to be necessary to estimate the model: consumption variables, income variables, and price variables. Starting with consumption variables, the adult household members will consume two private goods: food c_i and a non-food composite private good x_i. The BIHS contains a food consumption module that is going to allow the construction of an individual-level food consumption variable c_i. The module breaks down individual food consumption at each meal taken in the household in the 24 hours preceding the survey. Only meals in which the day laborer participates are kept to avoid issues with eating outside of the household. This module can be combined with data on food purchased by the household in the last week to get a precise measure of the food expenditures for each household member. The non-food private composite good x_i is constructed by summing expenditures on transport, communications, energy, hygiene, and cosmetics. This composite good is observed at the household level: $x_1 + x_2$ is observed, but not x_1 and x_2 separately. The adult household members also derive utility from a public good X: expenditures on children. This variable is constructed as the sum of expenditures on food consumed by children, clothes for children and education expenditures.

As noted above, the survey asks day laborers to report both their average weekly income (denoted $\bar{y}$) and their income over the past seven days (denoted y_s). From these two measures of income, the income shock $\tilde{y}$ is constructed as the ratio of the two: $\tilde{y} = \frac{y_s}{\bar{y}}$. Total household income Y is constructed as the sum of the day laborer's average weekly income and other income sources including wage income from other household members, income from selling agricultural products and remittances. Figure 1.1 shows histograms of the day laborer's two income measures, as well as $\ln(\tilde{y})$, which is the way in which the income shock will enter the empirical specification. The income shock $\ln(\tilde{y})$ is quite symmetrically distributed around 0. Note that the distribution of income in the last seven days is slightly to the right of the distribution of average weekly income. This is explained by the survey recording information on day laborers only if they had positive income in the past seven days. Most day laborers work in agriculture and are affected by seasonal availability of jobs. Since they are not necessarily working all year long, conditional on having worked in the past seven days, their income in the past seven days is higher than their average weekly income.

Figure 1.1: Day Laborer's Income



Notes: the figure on the left is a histogram of the day laborer's income in the past seven days (y_s) in blue and average weekly income ($\bar{y}$) with a black outline. The values are in Taka/day. 75 Taka $\approx$ 1 USD. The figure on the right is a histogram of the log-ratio of income in the last seven days and average weekly income $\ln(\tilde{y}) = (\frac{y_s}{\bar{y}})$.

While there is some sample selection, conditional on the probability of working in a given seven-day period, whether or not a day laborer worked in the seven days before the survey is random. Therefore, if two day laborers that have the same probability of working in a given seven-day period, whether they are included or not in the sample is random. In an effort to control for the unobserved probability of being in the sample, one of the empirical specifications will include an indicator variable for agricultural day laborers. The idea is that agricultural day laborers have more seasonal fluctuation in job availability and therefore are more likely to be affected by sample selection. The results, which are shown later, are unaffected by the inclusion of this indicator variable, which indicates that sample selection is not driving the results.

In addition, the timing of the survey, which determines the specific seven-day period over which income is reported, is a source of exogenous variation for the income shock. It is unlikely that interviewed day laborers would manipulate their income in the week before the interview or that the interviewers would systematically choose the date of the interview in any particular way. Therefore, if the interview date is random, the particular income draw over the past seven days should be random too. Of course, since we do not observe income draws of zero, the income draw is not actually random. However, conditional on the probability of being in the sample, the income shock can be thought of as random.

Table 1.2: Consumption, Income and Prices

	Mean	SD
Consumption		
Day Laborer Food (c_1)	39.7	17.0
Home Producer Food (c_2)	32.7	15.4
Non-food Private Good ($x_1 + x_2$)	37.3	16.4
Children's Consumption (X)	56.1	46.8
Total Expenditures (E)	169.7	73.9
Income		
Income in Last 7 Days (y_s)	147.3	77.8
Average Day Laborer Income ($\bar{y}$)	128.8	66.4
Total Household Income (Y)	209.1	132.7
Prices		
Food Price (p)	1.00	0.052
Children's Consumption Price (π)	1.00	0.045
Observations	1999	

Notes: Summary statistics (mean and standard deviation) describing consumption, income, and price variables for the sample that is used to estimate the model. All values are in Taka/day. 75 Taka $\approx$ 1 USD. The consumption variables are constructed by dividing expenditures by the respective price index. Income in the Last 7 Days and Average Weekly Income measure income of the day laborer. Total Household Income measures total average household income. Price indices are constructed at the village level.

The final variables required for estimation are prices. The price of the composite private good x_i is normalized to 1. A Laspeyres price index for food is constructed at the village level, using food items that are consumed by at least 10% of households. The weights used to construct the price index are average expenditure shares across the whole sample. A clothing price index is constructed similarly. The price of children's consumption is then a weighted average of the food price index and the clothing price index. The weights are the average share of children expenditures on food and clothing, respectively. Food expenditures and children expenditures are then divided by their respective price indexes to get the final consumption measures. Table 1.2 presents summary statistics for the variables discussed in this section.

1.5 Empirical Specification

Although the model is identified non-parametrically, precise estimates require reducing the dimensionality through a parametric specification. The parametric specification is similar to Cherchye, De Rock, Vermeulen (2012). The indirect utility from the consumption of food c_i and the non-food composite private good x_i, conditional on expenditures on children X is assumed to be:

$$V^i(p, \rho^i, X) = \frac{\ln(\rho^i) - \alpha^i \ln(p)}{p^{\beta^i}} + \kappa^i \ln(X),$$

where $\alpha^i = \alpha_0^i + \alpha_1^i K$ and $\kappa^i = \kappa_0^i + \kappa_1^i K$ with K the number of children in the household. In principle, α^i, β^i and κ^i could be functions of demographics. However, in order to avoid overspecification, only number of children is allowed to impact α^i and κ^i. As a reminder, p is the price of food consumption c_i and ρ^i is the budget allocated to agent i to spend on private goods. The preferences for the private good correspond to the preferences underlying Deaton and Muellbauer's (1980) Almost Ideal Demand System[9]. Using Roy's identity, the Marshallian demand for food consumption c_i can

[9]The notation α^i and β^i in this paper is consistent with the notation in Deaton and Muellbauer's (1980) seminal paper.

be derived.

$$c_i = [\alpha^i + \beta^i(\ln(\rho^i) - \alpha^i \ln(p))]\frac{\rho^i}{p} \tag{1.18}$$

Then, using the budget constraint from each agent's second stage, the demand for non-food consumption x_i can be derived.

$$x_i = \rho^i - pc_i = [(1-\alpha^i) - \beta^i(\ln(\rho^i) - \alpha^i \ln(p))]\rho^i \tag{1.19}$$

The demand for the public good can be derived from the first stage of the household's problem:

$$X = \kappa^1 \frac{\rho^1 p^{\beta^1}}{\pi} + \kappa^2 \frac{\rho^2 p^{\beta^2}}{\pi} \tag{1.20}$$

These demand functions depend on the sharing rule ρ^i. The sharing rule is determined in the first stage by the efficiency rule:

$$(\mu + \lambda_2)\frac{\partial V^1(p, \rho^1, X)}{\partial \rho^1} = (1-\mu)\frac{\partial V^2(p, \rho^2, X)}{\partial \rho^2}$$

Using the fact that $\frac{\partial V^1(p,\rho^i,X)}{\partial \rho^i} = \dfrac{1}{\rho^i p^{\beta^i}}$ and that $\rho = \rho^1 = y - \rho_2$ we get the following expression for the sharing rule ρ:

$$\rho = \frac{y}{1 + p^{(\beta_1 - \beta_2)}\dfrac{(1-\mu)}{(\mu+\lambda_2)}} \tag{1.21}$$

Following Browning, Chiappori and Lewbel (2008), the decision power of agent 1 for a given realization $\tilde{y}$ is assumed to take a logistic form:

$$\frac{\mu + \lambda_2}{1 + \lambda_2} = \frac{\exp(\gamma_1 \ln(\tilde{y}) + \gamma_2 \ln(\bar{y}) + \gamma_3 \ln(E) + \gamma_4 \ln(Y) + \gamma_D D)}{1 + \exp(\gamma_1 \ln(\tilde{y}) + \gamma_2 \ln(\bar{y}) + \gamma_3 \ln(E) + \gamma_4 \ln(Y) + \gamma_D D)} \tag{1.22}$$

Note that under the standard collective model, which assumes efficiency, the decision power should not vary in response to short-term income shocks. If the decision power varies with short-term income, this implies imperfect insurance in the household, which contradicts the efficiency assumption of the standard collective model. Therefore, under complete information, we would expect the decision power not to vary with income shocks $\tilde{y}$ and therefore we would have $\lambda_2 = 0$ and $\gamma_1 = 0$.

The decision power is allowed to depend on the wage earner's average income $\tilde{y}$. In general, the decision power can also vary across households. In order to capture some of that variation, the decision power can depend on a vector of demographic variables D and on total household income Y.

The decision power is also allowed to depend on total household expenditures E: the sum of all private and public expenditures in the household. This variable is included to capture potentially different risk aversions between the two adults in the household. For example, if the wage earner is less risk averse than the home producer, insurance within the household would imply that the wage earner's consumption share be greater when total household expenditures E are high. Short-term total household expenditures E and short-term income shock $\tilde{y}$ of the wage earner are likely to be correlated. Therefore, controlling for E is important to avoid a bias in $\hat{\gamma}_1$ and potentially overestimating the effect of asymmetric information.

The parametric specification of the decision power in equation (1.22) gives:

$$\frac{\mu + \lambda_2}{1 - \mu} = \exp(\gamma_1 \ln(\tilde{y}) + \gamma_2 \ln(\bar{y}) + \gamma_3 \ln(E) + \gamma_4 \ln(Y) + \gamma_D D)$$

Plugging this into equation (1.21) gives us an expression for ρ as a function of variables in the data:

$$\rho = y \times \frac{\exp((\beta_2 - \beta_1)\ln(p) + \gamma_1 \ln(\tilde{y}) + \gamma_2 \ln(\bar{y}) + \gamma_3 \ln(E) + \gamma_4 \ln(Y) + \gamma_D D)}{1 + \exp((\beta_2 - \beta_1)\ln(p) + \gamma_1 \ln(\tilde{y}) + \gamma_2 \ln(\bar{y}) + \gamma_3 \ln(E) + \gamma_4 \ln(Y) + \gamma_D D)} \tag{1.23}$$

Finally, this equation for ρ can be plugged into the demand equations (1.18), (1.19), and (1.20) to get closed-form solutions for demands as functions of data and parameters.

This system of equations can then be estimated. To be more precise, we will model $(c_1, c_2, x = x_1 + x_2, X)$ as observable functions of $(p, \pi, y, \tilde{y}, \bar{y}, E, Y, D)$. Additive errors are added to the demand equations to account for unobservable heterogeneity. The system of equations is estimated with a feasible generalized nonlinear least squares estimator. Note that this estimator allows for correlated errors across the different goods.

1.6 Results and Counterfactuals

1.6.1 Main Result: the Effect of Asymmetric Information

Since complete information is a special case of the model, the estimation results provide a natural test for asymmetric information. Under the null hypothesis of complete information, the effect of the income shock should be zero. Therefore, if the effect of an income shock on the decision power is positive and statistically significant, we can reject complete information.

The main estimation results are given in Table 1.3. The first and second columns correspond to the estimated coefficients and corresponding standard errors when only total household income Y is included in the decision power. The third and fourth columns correspond to the case where two other demographic variables are included: an indicator variable for whether the day laborer works in agriculture and the number of children in the household.

The first thing to note is that the estimated coefficient of the income shock $\tilde{y}$ is, in fact, positive and statistically significant. This allows us to reject complete information and provides evidence that these households are affected by asymmetric information. The last row of the table quantifies the effect of an income shock: in the baseline specification, a one standard deviation income shock above the mean increases the consumption share of the wage earner by 0.49 percentage points.

The baseline specification and the specification with the additional demographic

Table 1.3: Structural Estimation Results

	Baseline specification		Demographic Controls	
	Coef.	SE	Coef.	SE
Decision Power				
γ_1[Income Shock]	0.030*	0.0099	0.031*	0.0098
γ_2[Average Income]	0.031*	0.0089	0.031*	0.0089
γ_3[Tot. Expenditures]	-0.20*	0.014	-0.21*	0.014
γ_4[Total HH Income]	0.34*	0.015	0.32*	0.016
γ_{D1}[Works in Agr.]			0.011	0.0095
γ_{D2}[Number of Kids]			0.12*	0.022
Private Preferences				
α_0^1[Constant]	0.84*	0.032	0.82*	0.032
α_1^1[Number of Kids]	-0.011*	0.0017	-0.023*	0.0028
β^1	-0.060*	0.0046	-0.052*	0.0044
α_0^2[Constant]	0.15*	0.068	-0.035	0.093
α_1^2[Number of Kids]	-0.053*	0.0049	0.059*	0.025
β^2	0.26*	0.037	0.27*	0.039
Public Preferences				
κ_0^1[Constant]	0.0061	0.078	0.059	0.065
κ_1^1[Number of Kids]	0.18*	0.030	0.15*	0.022
κ_0^2[Constant]	-0.068	0.19	-0.35*	0.17
κ_1^2[Number of Kids]	0.36*	0.070	0.54*	0.065
N	1999		1999	
1 SD Income Shock	0.49 p.p.		0.48 p.p.	

Notes: The model is estimated using a feasible generalized nonlinear least squares estimator. * indicates that the coefficient is significant at the 5% level. Columns 1 and 2 give the results (estimated coefficient and asymptotic standard errors) for the baseline specification without demographic controls in the decision power. Columns 3 and 4 give the results for the estimation with demographic controls (a dummy for whether the day laborer works in agriculture and the number of children) in the decision power. The expressions in brackets refer to the objects that are related to the respective parameters (see equations 18-23). The last row indicates by how much the consumption share increases in response to an income shock one standard deviation above the mean.

controls are extremely similar for most coefficients. In particular, the addition of these demographic controls does not affect how the decision power varies with the income shock. The fact that the estimation is not sensitive to the inclusion of demographic controls provides some evidence that the results are not being driven by unobserved heterogeneity across households.

The estimation of the preference parameters suggest that wage earners and home producers have different preferences. The wage earner prioritizes food consumption, as can be seen by the large α_0^1 relative to α_0^2. The home producer also has a stronger preference for expenditures on children, consistent with previous results in the literature.

1.6.2 Work Location Heterogeneity

Some households are likely to be more affected by asymmetric information than others. In particular, if the wage earner works farther from home, income shocks are likely to be less observable to other household members. Then, in order to be incentivized to report truthfully, such a wage earner's consumption share would be more responsive to income shocks.

Analyzing heterogeneity in the response of consumption shares to income shocks serves two main purposes. First, it provides an additional test for whether the asymmetric information mechanism is driving the main results. The prediction under asymmetric information is that the response of consumption shares to income shocks would be larger for wage earners working far from home relative to those working close to home. The heterogeneity analysis tests this prediction. Second, identifying households that are likely to be more affected by asymmetric information could help target public policies aimed at reducing the effect of asymmetric information.

The work location heterogeneity is modeled by allowing the effect of income shocks on the decision power to be different for two groups: those that work far from their home and those that work near their home. Specifically, 11% of wage earners work outside of the union in which they live[10]. The model is reestimated allowing the effect of income shocks on decision power to be different for these wage earners.

More formally, d is defined to be equal to 1 if the wage earner works outside of the union in which they live, 0 otherwise. The main empirical specification is augmented by having $\gamma_1 = \gamma_{10} + \gamma_{11} \times d$ and including d additively in the decision power equation. This is the only change to the model in the baseline heterogeneity analysis. In a second analysis, in addition to the baseline changes, the effect of other variables on decision power is also allowed to differ for the two subsamples by having $\gamma_i = \gamma_{i0} + \gamma_{i1} \times d$ for $i = 2, 3, 4$.

The results are presented in Table 1.4. The first two columns correspond to the case

[10]Unions are the smallest administrative unit in Bangladesh. There are 4562 unions in Bangladesh, with an average area of 32 square kilometers.

Table 1.4: Work Near Home vs. Work Far From Home

	Shock Interaction		All Interactions	
	Coef.	SE	Coef.	SE
Work Near ($d = 0$)				
γ_{10}[Income Shock]	0.018	0.0100	0.022*	0.010
γ_{20}[Average Income]	-0.0038	0.011	-0.013	0.012
γ_{30}[Tot. Expenditures]	-0.20*	0.013	-0.19*	0.014
γ_{40}[Total HH Income]	0.35*	0.015	0.36*	0.015
Work Far ($d = 1$)				
γ_{11}[Income Shock]	0.10*	0.022	0.081*	0.033
γ_{21}[Average Income]			0.045	0.031
γ_{31}[Tot. Expenditures]			-0.055	0.038
γ_{41}[Total HH Income]			-0.048	0.033
γ_{01}[Constant]	-0.45*	0.11	-0.024	0.22
N	1999		1999	

Notes: The model is estimated using a feasible generalized nonlinear least squares estimator. * indicates that the coefficient is significant at the 5% level. Columns 1 and 2 give the results (estimated coefficient and asymptotic standard errors) when only the income shock variable is interacted with the dummy for work location. Columns 3 and 4 give the results when all variables in the decision power are interacted with the dummy. The expressions in brackets refer to the objects that are related to the respective parameters (see equation 23). The top panel (above the dashed line) is the parameter value for those that work near home. The parameter values for those that work far from home is the sum of the value in the top panel and the associated value in the bottom panel (e.g. $\gamma_{10} + \gamma_{11}$). 210 day laborers work far from home and 1789 near home.

where only the effect of the income shock is allowed to vary across the two subsamples. The second pair of columns correspond to the case where all the variables are interacted.

The results provide strong evidence in favor of the asymmetric information mechanism presented in this paper. Consistent with the effect of asymmetric information, the wage earners' consumption increases significantly more with income shocks for those that work far from their home. In fact, the magnitude of the effect is 4 to 5 times larger, depending on the specification. In addition, note that the effect of the other variables on decision is not statistically different for the two subsamples of wage earners. This helps address the concern that the difference in the effect of income shocks is being driven by selection or preference heterogeneity.

The large difference in the effect of income shocks for the two subsamples suggests that households are affected quite heterogeneously by asymmetric information. As suggested by the results in table 2, this heterogeneity is likely to be the result of variation in how observable income shocks are to other household members. In addition, variation in altruism or preferences for hiding income might also be important dimensions of heterogeneity to consider. Overall, this suggests the importance of targeting policies to reduce asymmetric information appropriately.

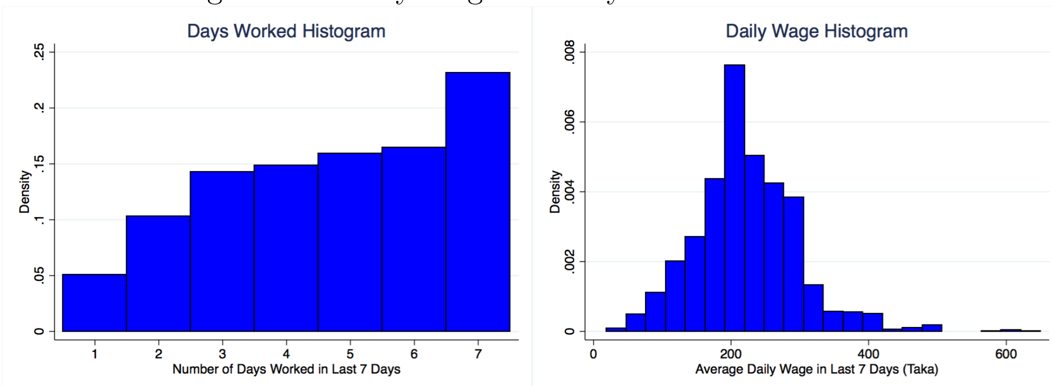
1.6.3 Daily Wage and Days Worked Variation

A possible alternative explanation for the previous empirical results is that consumption shares and income shocks are correlated because more resources are needed by the wage earner when working more. This is particularly a concern because the assignable good being used to identify the decision power is food consumption. Since most of the wage earners work in manual labor industries, it is likely that when they work more they have higher nutritional needs.

In order to show that nutritional needs are not the main explanation for the empirical results, income variation is separated between daily wage and days worked variation. The prediction is that if nutritional needs are driving the results, the variation in consumption shares should be explained by the variation in days worked. If, on the contrary, asymmetric information is driving the results, the variation in consumption shares should be explained by the variation in daily wage. The reasoning is that nutritional needs of a worker will depend more on the number of days worked than on the daily wage. In addition, daily wage variation is likely to be less observable by other household members than how many days the wage earner worked. Comparing the effect of wage variation and days worked variation could therefore serve as a test of the asymmetric information mechanism, even in the absence of nutritional needs. Here, the comparison serves the added purpose of ruling out an alternative mechanism.

Separating daily wage and days worked variation requires observing measures of these variations. The measure of the short-term income shock that has been used up to now is the ratio of income over the last seven days (y_s) and average weekly income ($\bar{y}$): $\tilde{y} = \frac{y_s}{\bar{y}}$. Income over the last seven days was constructed from the survey data as the product of number of days worked in the last seven days (y_d) and average daily wage in the last seven days (y_w): $y_s = y_d \times y_w$. Since the income shock enters the decision power expression through a logarithmic function, it is natural to divide the wage and

Figure 1.2: Daily Wage and Days Worked Variation



Notes: the figure on the left is a histogram of the number of days worked in the last seven days by the day laborer. The figure on the right is a histogram of the day laborer's average daily wage over the last seven days. The daily wage is in Taka. 75 Taka ≈ 1 USD.

days variation using:

$$\ln(\tilde{y}) = \ln(\frac{y_w \times y_d}{\bar{y}}) = \ln(y_w) + \ln(y_d) - \ln(\bar{y})$$

Then we can reestimate the model by including $\gamma_{1w} \ln(y_w) + \gamma_{1d} \ln(y_d)$ in the expression for the decision power instead of just $\gamma_1 \ln(\tilde{y})$ [11]. If the results are not being driven by nutritional needs, we would expect a large and positive γ_{1w}.

Figure 1.2 plots the histograms of the two sources of income variation. The main takeaway is that days worked varies quite a bit and does not seem to follow the usual five-day work week that is common in developed countries. This is an important point because without significant days worked variation it would be impossible to compare the effect of the two sources of income variation.

The estimation results for the decision power are given in Table 1.5. They rule out nutritional needs as the driver of the empirical results in the main specification. Consistent with the asymmetric information mechanism, the coefficient on daily wage is positive and significant. In fact, the estimated coefficient for daily wage is twice as large as the one for days worked. The last two rows allow a more intuitive comparison

[11] Since $\gamma_2 \ln(\bar{y})$ is already in the main specification, there is no need to include another term for it. Simply note that in the new estimation the coefficient on $\ln(\bar{y})$ will correspond to $\gamma_2 - \gamma_1$ from the original specification.

Table 1.5: Daily Wage vs. Days Worked Variation

	Coefficient	Standard Error
γ_{1w}[Daily Wage]	0.050*	0.015
γ_{1d}[Days Worked]	0.022*	0.011
γ_2[Average Income]	-0.00029	0.011
γ_3[Tot. Expenditures]	-0.20*	0.014
γ_4[Total HH Income]	0.34*	0.016
N	1999	
$\gamma_{1w} - \gamma_{1d}$ (pval)	0.28	0.069
1 SD Shock (Wage)	0.33 p.p.	
1 SD Shock (Days)	0.17 p.p.	

Notes: The model is estimated using a feasible generalized nonlinear least squares estimator. * indicates that the coefficient is significant at the 5% level. Columns 1 and 2 give the results (estimated coefficient and asymptotic standard errors). The expressions in brackets refer to the objects that are related to the respective parameters. The seventh row reports the difference between the parameter on the daily wage and the parameter on the number of days worked, and the associated p-value. The second-to-last row indicates by how much the consumption share increases in response to a one standard deviation increase in the daily wage. The last row indicates by how much the consumption share increases in response to a one standard deviation increase in days worked.

of the magnitudes. A daily wage one standard deviation above the mean leads to a 0.33 percentage point increase in the wage earner's consumption share. Similarly, a one standard deviation increase in the days worked leads to a 0.17 percentage point increase in the wage earner's consumption share. This clearly shows that the variation in consumption share in response to income shocks results from daily wage variation, more so than from days worked variation. Therefore, the analysis is consistent with asymmetric information as the main mechanism, rather than nutritional needs.

1.6.4 First Counterfactual: A Naive Approach to Asymmetric Information

The model presented in this paper emphasizes that cooperative households can reduce the cost of asymmetric information by incentivizing wage earners to reveal their true income. The empirical results provide evidence that the households do in fact behave in this way. However, this raises the question: what would be the welfare cost for households if they did not incentivize truthful income reporting? This question is important because it can help shed light on the experimental results on asymmetric information. Experiments are typically one-offs and situations household members have no experience dealing with. As a result, household members are likely not able to coordinate to provide incentives in an experiment as they would in a repeated setting outside of the lab. Therefore, experiments might overstate the cost of asymmetric information.

Wage earners that are not incentivized will choose to hide some income. How much will depend on their preferences for hiding income. Therefore, quantifying the welfare cost for a "naive" household of not incentivizing the wage earner requires estimating preferences for hiding income. The identification result states that these preferences can be estimated if a hideable good is observed at the individual level. A hideable good is simply a good that can be consumed without other household members' knowledge. Here the hideable good will be betel nuts, a widely used stimulant in Bangladesh, similar to tobacco. Individual level consumption of betel nuts is observed. 671 out of the 1999 day laborers consume betels nuts and therefore the counterfactual analysis will focus on this subsample.

The counterfactual simulation then proceeds in two steps. In the first step, preferences for betel nuts (denoted h_1) are estimated. The parametric specification is chosen to be $v^1(h_1) = \tau_1 \ln(h_1)$. A closed form solution for h_1 as a function of parameters can be derived by combining the first order condition for the hideable good (1.24) and the partial derivative of the incentive compatibility constraint from the first stage of the household's problem with respect to $\tilde{y}$ (1.25):

$$v_1'(h_1) = \frac{\partial V^1(p, \rho^1, X)}{\partial \rho^1} + \frac{\lambda_2}{\mu + \lambda_2} \frac{1}{\frac{\partial h_1}{\partial \tilde{y}}} v_1'(h_1) \tag{1.24}$$

$$v_1'(h_1) = \frac{\partial V^1(p, \rho^1, X)}{\partial \rho^1} \frac{\partial \rho^1}{\partial \tilde{y}} + \frac{\partial h_1}{\partial \tilde{y}} v_1'(h_1) \tag{1.25}$$

The choice of functional forms for the utility functions gives us:

$$h_1 = \tau_1 \frac{(\gamma_1 + \tilde{y}) - \sqrt{(\gamma_1 + \tilde{y})^2 - 4\frac{\mu}{\mu + \lambda_2}\gamma_1 \tilde{y}}}{2\gamma_1 \tilde{y}}$$

Estimating τ_1 using this equation would require knowing γ_1 and the other parameters that go into μ and λ_2. Obviously, the true values are unknown. However, since these parameters have already been consistently estimated in the main estimation results, the estimated values can be plugged in instead. τ_1 can then be consistently estimated using a linear least squares estimator. However, the two-stage estimation means that the standard errors for τ_1 have to be bootstrapped[12].

The second step of the counterfactual is to simulate the behavior of wage earners that are not incentivized to reveal their income. Without incentives, the wage earner will decide how much to hide based on the preference for hiding income relative to revealing income. This choice is determined by the following first-order condition:

$$v_1'(h_1) = \frac{\partial V_1(p, \tilde{\rho}^1, X)}{\partial \tilde{\rho}^1} \frac{\partial \tilde{\rho}^1}{\partial y}$$

$\tilde{\rho}^1$ is the sharing rule in this naive scenario, where the within-household consumption share does not vary with reported income. Therefore, compared to equation (1.23), the only difference with this modified sharing rule is that it imposes $\gamma_1 = 0$. The intuition is that the wage earner will choose to hide until the marginal utility from hiding is equal to the marginal utility from reporting truthfully. The marginal utility from reporting truthfully is equal to the product of the marginal increase in the wage earner's share of resources and the marginal utility increase from the extra resources. Given the functional forms, this gives us:

$$h_1^c = \frac{p_1^\beta y}{\tau_1}$$

where h_1^c is the counterfactual value of the betel nut consumption. This equation determines the maximum the wage earner would like to hide. In order to have a more realistic simulation, the wage earner is constrained to hiding a maximum of 10% of income over the past seven days in the baseline simulation and 20% in a second simulation. This reflects the fact that only part of the income is unobservable to other household members and can be hidden.

The results are presented in Table 1.6. When wage earners can hide up to 10%

[12] I have not finalized the bootstrap. For now I do not report standard errors.

of income, they hide on average 44% of that 10%. 15% choose to hide the maximum amount they can hide. When they can hide up to 20% of income, they hide on average 26% of that 20%, and 4.3% choose to hide the maximum amount they can hide.

Table 1.6: Counterfactual Analysis Results

Unobservable Income	τ_1	% Hidden	Share hiding all
10%	0.11	44%	15.1%
20%	0.11	26%	4.3%

Notes: The rows correspond to a simulation where 10% and 20% of income are not observed by the spouse, for the first and second row respectively. The first column is the estimated coefficient for τ_1 - the preference parameter for the hideable good. The second column gives the average share of unobservable income actually hidden. The third column gives the share of day laborers hiding all the unobservable income.

Interestingly, wage earners choose to report a large portion of the income they could have hidden, even in the absence of incentives. This reflects a preference for goods that are consumed in the household, such as food and other essentials. It is also consistent with the relative small consumption share changes that were estimated in the main specification: a 0.49 percentage point increase in wage earners' consumption share is enough for them to report an income shock one standard deviation above the mean.

The welfare implications of not incentivizing truthful income reporting are concentrated on the spouse and children. The counterfactual shows that when 10% of income can be hidden, incentives increase the welfare of the spouse and children equivalently to a 4.4% increase in total household income. When the wage earner can hide up to 20% of income, that number goes up to 5.2%. Meanwhile, the wage earner experiences a welfare gain equivalent to a 0.2% increase in household income[13]. Therefore, the welfare gains from providing incentives are considerable.

1.6.5 Policy Counterfactuals

Given the estimation of the welfare functions, it is possible to simulate counterfactual policies and quantify their effect. Two policies will be considered in this part: a guaranteed employment scheme and a tax on hideable consumption.

[13]While the IC constraint constrains the wage earner to not be better off by hiding, here there are very slight welfare gains because the amount of betel nut consumed is chosen optimally through the FOC rather than being observed in the data.

First, I consider a guaranteed employment scheme. This policy is used in developing countries, such as Bangladesh and India, to help poor workers through periods of low employment. The daily wage of participants is constant and fixed nationally, and therefore is not likely to be affected by asymmetric information. Therefore, a day laborer working in a guaranteed employment scheme does not need to be incentivized to report their income. The policy simulation therefore consists in shutting down the asymmetric information mechanism by setting $\gamma_1 = 0$. I then compute the compensating variation in income: by how much would total household income have to decrease for the household to be indifferent between, on the one hand, the policy and the lower income and, on the other, the pre-policy situation.

Quantifying the welfare gains from this policy requires choosing an increasing transformation for the welfare functions. The reason is that this policy increases welfare through better risk sharing within the household. Therefore, the welfare gains depend on the degree of concavity of the welfare functions. Unfortunately, the identification result tells us that the welfare functions are only identified up to an increasing transformation. The empirical specification chosen to estimate the model implies a constant relative risk aversion of 1. This is well within the range of risk aversion estimates found in previous work. In fact, looking at Cardenas and Carpenter's 2008 review of the literature, a constant relative risk aversion coefficient of 1 would be pretty close to the median estimate. Therefore, the increasing transformation chosen for the welfare analysis will simply be the identity function. In other words, the welfare functions are kept unchanged.

I find that shutting down the asymmetric information channel through a guaranteed employment scheme has modest welfare effects. Using equal weights on the welfare functions of the day laborer and the home producer, I find that, on average, households are willing to pay 0.3% of total yearly household income to remove asymmetric information through a guaranteed employment scheme, for a one year period. This small effect is explained by the relatively small effect of short-term income shocks on consumption shares: a one standard deviation short-term income shock increases the day laborer's consumption share by only 0.49 percentage points. Therefore, the welfare gain from shutting down this inefficient "extra" variation is small.

Second, I consider a tax on hideable consumption. Just as in the naive approach counterfactual, I use betel nuts as the hideable good, which restricts my sample to the 671 day laborers that consume it. I consider the effect of a 50% tax on betel nuts, which results in a 50% increase in the price of betel nuts. This policy has two main welfare effects. First, it decreases the benefit of hiding income for day laborers since they will be able to buy less with it. Therefore, day laborers do not need to be incentivized as much to report truthfully. As a result, consumption share variation due to asymmetric information will decrease, resulting in welfare gains through better insurance within the household. Second, the policy increases the cost of a good that is valued by household members, which will decrease welfare.

I then simulate the policy and compute welfare functions. I find that the policy is a net welfare loss for the household. The average willingness-to-pay for the policy is actually -2.7%. The welfare cost - an increased price for a valued good - actually outweighs the benefit in terms of reduced consumption share variance. However, the policy does reduce the "extra" variance in consumption share from 26%to 16% of the total variance in consumption share.

1.7 Conclusion

This paper provides evidence that asymmetric information about income matters for household decisions, resource-sharing and welfare. A new model of a household facing asymmetric information about the income of the wage earner provides a theoretical framework to predict the effect of asymmetric information on household outcomes and to conduct counterfactual analyses. Then, a new identification result, specific to this model, states that individual welfare functions can be estimated from consumption data. Finally, the model is estimated using a sample of Bangladeshi day laborers. The estimation provides evidence that the day laborers are affected by asymmetric information but that cooperative households are able to significantly reduce the cost of asymmetric information by providing incentives.

While this paper finds relatively small welfare effects of asymmetric information, there is reason to believe that in some specific contexts the effects of asymmetric information might be more dramatic. In particular, the much larger effect of asymmetric information for the workers that work far from home points to a high degree of heterogeneity. At least in the Bangladeshi context, households differ greatly in how affected they are by asymmetric information. Therefore, it would be important to understand what drives this heterogeneity. This could help target policies aimed at reducing the welfare cost of asymmetric information. One situation in which asymmetric information might be particularly important is when cooperation breaks down in the household. For example, divorce law in the United States requires spouses to disclose all assets. In this context the ability of one spouse to hide income might have important consequences on the divorce decisions and on how resources are shared following the divorce. This question is beyond the scope of this paper but merits further investigation.

Appendix

We will show that the knowledge of the ratio of partial derivatives of a function allows us to recover that function up to an increasing transformation. We will show this for a function with three arguments because this is the relevant case in this paper. However, this argument can be generalized to an arbitrary function.

We are interested in the function $W(x, y, z)$. Suppose that the function is monotonic in each one of its arguments. If it is not, we can apply the argument below over monotonic regions of the function. Denote W_ξ for $\xi = x, y, z$ the partial derivatives of this function. We know $r^1(x, y, z) = W_y/W_x$ and $r^2(x, y, z) = W_z/W_x$ the ratios of partial derivatives. The question is: if two functions $W(x, y, z)$ and $\hat{W}(x, y, z)$ have the same ratios of partial derivatives, what is the relationship between of $W(x, y, z)$ and $\hat{W}(x, y, z)$? The goal is to show that it must be that $\hat{W}(x, y, z) = G(W(x, y, z))$ for some increasing function $G(.)$.

First note that $W_y/W_x = \hat{W}_y/\hat{W}_x$ and $W_z/W_x = \hat{W}_z/\hat{W}_x$ imply $\hat{W}_x/W_x = \hat{W}_y/W_y = \hat{W}_z/W_z \equiv g(x, y, z)$ for some function $g(x, y, z)$. Now we have that $\hat{W}_\xi = g(x, y, z)W_\xi$ for $\xi = x, y, z$.

The next step is to integrate the partial derivatives of $\hat{W}(x, y, z)$:

$$\hat{W}(x, y, z) = \int \hat{W}_x(x, y, z)dx = \int g(x, y, z)W_x(x, y, z)dx$$

Then apply the implicit function theorem to the relationship $W(x, y, z) = W$ to define the function $h(y, z, W)$ such that $h(y, z, W(x, y, z)) = x$. Then we can write the equation above as:

$$\hat{W}(x, y, z) = \int g(h(y, z, W(x, y, z)), y, z)W_x(x, y, z)dx$$

Then we can integrate by substitution to find:

$$\hat{W}(x, y, z) = \int g(h(y, z, W), y, z)dW = H^1(y, z, W(x, y, z)) + k^1(y, z) \tag{A1}$$

Simlarly, we find:

$$\hat{W}(x, y, z) = H^2(x, z, W(x, y, z)) + k^2(x, z) \tag{A2}$$
$$\hat{W}(x, y, z) = H^3(x, y, W(x, y, z)) + k^3(x, y) \tag{A3}$$

Then we can take the partial derivatives of (A1), (A2), (A3) with respect to x, y, z to find the following relationships:

$$\begin{aligned}
\hat{W}_x &= g(x, y, z)W_x = H_3^1 W_x = H_3^2 W_x + H_1^2 + k_1^2 = H_3^3 W_x + H_1^3 + k_1^3 \\
\hat{W}_y &= g(x, y, z)W_y = H_3^2 W_y = H_3^1 W_y + H_1^1 + k_1^1 = H_3^3 W_y + H_2^3 + k_2^3 \\
\hat{W}_z &= g(x, y, z)W_z = H_3^3 W_z = H_3^1 W_z + H_2^1 + k_2^1 = H_3^2 W_z + H_2^2 + k_2^2
\end{aligned}$$

The first thing to notice is that $H_3^1 = H_3^2 = H_3^3 = g(x, y, z)$. This, in turn, implies that $H_i^j = -k_i^j$ for $i = 1, 2$ and $j = 1, 2, 3$. Notice that the k functions depend on only two arguments, while the H functions depend on all three. Therefore, we find that the partial derivative of the H functions with respect to the first two arguments do not depend on the variable that enters the H functions only through W. For example $H_1^1(x, y, z) = -k_1^1(y, z)$ and $H_2^1(x, y, z) = -k_2^1(y, z)$. Therefore, the partial derivatives of H^1 with respect to the first two arguments do not depend on x. This, in turn implies the function H^1 is additively separable between the first two and the third argument. In other words, the function takes the form:

$$H^1(y, z, W(x, y, z)) = \tilde{H}(W(x, y, z)) + l^1(y, z)$$

In addition, we also get that $H^1_1(y, z, W(x, y, z)) = l^1_1(y, z) = -k^1_1(y, z)$ and $H^1_2(y, z, W(x, y, z)) = l^1_2(y, z) = -k^1_2(y, z)$. Therefore, the sum $l^1(y, z) + k^1(y, z)$ is a constant that does not depend on y or z. Putting this into equation (A1) we find that $\hat{W}(x, y, z) = \tilde{H}(W(x, y, z)) + k = G(W(x, y, z))$ for some function $G(.)$. Therefore, we have shown that the ratios of partial derivatives allow us to recover the function up to some transformation.

In addition, if we know the sign of one of the partial derivatives of W, say for example $W_x > 0$, then this implies that $G(.)$ has to be an increasing function. This would be the case in this paper because it makes sense to assume that preferences are such that collective utilities are decreasing in p, for example.

CHAPTER 2

Individual Rather Than Household Euler Equations: Identification and Estimation of Individual Preferences Using Household Data

Maria Casanova, Alexandre Fon, Maurizio Mazzocco [1]

2.1 Introduction

The evaluation of many public policies requires knowledge of the preferences that determine the behavior of multiperson households. Changes in tax rates on pension assets, asset-based means-tested welfare programs, and marriage penalty relief programs are only a few examples. The traditional approach for estimating preferences assumes that households behave as single agents. Under this assumption, each household can be characterized using a unique utility function independently of the household structure. Since the unique utility function depends on household total consumption, which is

[1]Maurizio Mazzocco thanks the National Science Foundation for Grant SES-0231560. We are very grateful to Orazio Attanasio, Pierre André Chiappori, Lucas Davis, James Heckman, John Kennan, and Annamaria Lusardi for their insight and suggestions. We would also like to thank participants at SITE, Econometric Society Summer Meeting, and seminars at CREST-Paris, Duke University, Stanford University, University of California at Berkeley, University of Maryland, University of Minnesota, University of North Carolina, University of Wisconsin-Madison, for their invaluable comments. Errors are ours. Maria Casanova, Cal State Fullerton, Department of Economics, Fullerton, CA. Email: mcasanova@exchange.fullerton.edu. Alexandre Fon, UCLA, Department of Economics, Los Angeles, CA. Email: afon@ucla.edu. Maurizio Mazzocco UCLA, Department of Economics, Los Angeles, CA. Email: mmazzocc@econ.ucla.edu.

observed, the intratemporal and intertemporal features of household preferences can be identified and estimated using standard methods.

Numerous papers have rejected the hypothesis that households behave as single agents. For instance, the results of Schultz (1990), Thomas (1990), Browning et al. (1994), Browning and Chiappori (1998), and Mazzocco (2007) indicate that micro-level data are not consistent with this hypothesis. The main implication of this finding is that public policies cannot be evaluated using a unique household utility function, since as shown in Mazzocco (2004) important aspects of intra-household risk sharing and specialization are ignored. Estimates of the preferences of each decision maker in the household are required. The main obstacle in the identification and estimation of individual preferences is that they depend on individual consumption, which is generally not observed. The goal of this paper is to identify and estimate such preferences using the limited amount of information which is available in household surveys. This is one of the first attempts to identify and estimate the intertemporal features of individual preferences by taking into account that household-level data are the outcome of joint decisions by household members.

This paper makes two main contributions. First, it is shown that the preferences of each decision maker in the household can be identified even if individual consumption is not observed, provided that household consumption, individual labor supply, and individual wages are observed. To illustrate the idea behind this result, consider a married couple. If individual consumption were observed, individual preferences could be identified by standard methods using individual Euler equations, i.e. one set of intertemporal optimality conditions for each agent, and intraperiod optimality conditions. Individual consumption is generally not observed, but household consumption, individual labor supply, and wages provide information on this variable. In particular, if at least one agent works in each period, the marginal rate of substitution between individual consumption and leisure should equal the real wage. As a consequence, this agent's consumption can be written as a function of labor supply and wages. Since consumption of the second agent is equal to the difference between total household consumption and consumption of the first agent, the spouse's consumption can also be written as a function of observed variables. These functions can be used to substitute out individual consumption from the marginal utilities that define the individual Euler equations and intraperiod optimality conditions. It can then be shown that these reduced-form optimality conditions and variations in household consumption, individual labor supply, and wages provide sufficient information to recover the original utility functions.

As a second contribution, individual preferences are estimated using the described identification result, a specific functional form for the individual utility functions, and data from the CEX. To evaluate the performance of the identification result, individual preferences are first estimated for single females and males with no children. For this group of households, individual consumption is observed since it is equivalent to household consumption. Individual preferences can therefore be estimated using the identification method proposed in this paper as well as standard methods. The results indicate that the identification method performs well in the sense that the parameter estimates obtained using the identification result are comparable to the estimates obtained using standard methods. The empirical findings also suggest that there is heterogeneity in intertemporal preferences between single females and single males: the intertemporal elasticity of substitution of single males is more than twice the corresponding elasticity for single females.

The identification result is then applied to a sample of couples. Similarly to single individuals, we find strong evidence of heterogeneity in intertemporal preferences between wives and husbands. In particular, the intertemporal elasticity of substitution of wives is about half the elasticity for husbands or, equivalently in this paper, wives are about twice as risk averse as husbands. A comparison of the parameter estimates for single and married agents indicates that single males are less risk averse than married males and that single females are more risk averse than married females.

These findings have one main implication. In Mazzocco (2007), it is shown that households behave as single agents only if individual preferences belong to the Harmonic Absolute Risk Aversion (HARA) class with identical curvature parameter. The preference heterogeneity found in this paper indicates that this condition is not satisfied. Therefore economists and policy makers should not rely on preference estimates obtained using the standard unitary model to evaluate alternative policy recommendations. Instead, policy analysis should be performed using individual preferences and the corresponding parameter estimates.

This paper is related to the literature on the collective representation of household behavior. Manser and Brown (1980) and McElroy and Horney (1981) are the first papers to characterize the household as a group of agents making joint decisions. In those papers the household decision process is modeled as a Nash bargaining problem. Chiappori (1988; 1992) extends their analysis to allow for any type of efficient decision process. The theoretical model used in the present paper is an intertemporal generalization of Chiappori's collective model.

The intraperiod features of individual preferences have been identified and estimated in other papers. For instance, Blundell et al. (2005), Blundell et al. (2007), Donni (2009), Chiappori (1988; 1982), Fong and Zhang (2001) show that different aspects of intraperiod preferences can be identified. Donni (2009) also estimates them. The present paper is, however, one of the first attempts to identify and estimate the intertemporal features of individual preferences using household data.

The chapter is organized as follows. In section 2, the individual Euler equations are derived. Section 3 outlines the identification procedure. Section 4 describes the empirical implementation. Section 5 discusses econometric issues. Section 6 describes the data and section 7 presents the estimation results. Section 8 concludes.

2.2 Euler Equations of Singles and Couples

In this section, we test the hypothesis that the constant rejection of household Euler equations is generated by the fact that most of the households included in their estimation are composed of two or more decision makers who make joint decisions based on their individual decision power. The section will be divided into three parts. In the first subsection, we provide theoretical arguments that explain why the inclusion of married and cohabiting couples in the estimation of Euler equations will generally result in their rejection. We then introduce a test that enables us to evaluate whether our hypothesis can explain the rejections of the household Euler equations. In the second subsection, we describe the data that will be used in the implementation of the test. In the last subsection, we discuss how the tests is implemented, some econometric issues, and our empirical results.

2.2.1 Rejections of Euler equations and Household Composition

In the past three decades, Euler equations have been extensively used to test intertemporal models and to estimated preference parameters. The most frequently used test that relies on Euler equations is the excess sensitivity test proposed by Hall (1978) and Sargent (1978), which is based on the following idea. Household should choose current and future consumption according to the Euler equations using all the information available at the time of the decision. As a consequence, the difference between the current marginal utility of consumption and next period expected marginal utility of consumption should be independent of variables that are known to the household at the time of the decision. Many papers have tested this implication of the standard intertemporal model and regularly rejected it.

The model generally used to derive Euler equations, the unitary model, is well-suited to characterize the intertemporal decisions of households composed of a single individual. It is therefore also a good model to estimate the intertemporal dimension of the preferences of single individuals, such as their risk aversion or intertemporal elasticity of substitution. That model is not, however, well-suited to represent the behavior of married or cohabiting couples for reasons that will be discussed in the next paragraph. In spite of this, papers that estimate Euler equations rely either on a sample composed exclusively of couples, which is the case for instance in Attanasio and Browning (1995) and Meghir and Weber (1996), or alternatively on a sample that includes both couples and single individuals, as it is the case for example in Attanasio and Weber (1995) and Zeldes (1989). A potential explanation for the constant rejections of the Euler equations is therefore that the model used to derive them is not the proper model for most of the households included in the estimation.

To understand why the model employed to derive Euler equations is problematic when used to characterize couples, we will introduce a generalization to an intertemporal setting of the collective model of the household, which is by now a standard framework to study household decisions. The main feature of the collective model is that it explicitly recognize that the majority of households are composed of several individuals with potentially heterogeneous preferences who make joint decisions. This is done by assigning a utility function to each household member and by aggregating the individual preferences using the assumption that decisions are Pareto efficient. We will denote with u^i the utility function of member i and with μ the Pareto weight. Since in the U.S. most households with several decision makers are couples, we will consider the case in which households are composed of two members. To be consistent with the literature on the estimation of Euler equations we will assume that individuals have preferences over non-durable consumption c. Specifically, we assume that preferences are strongly separable between consumption and leisure and that there is no distinction between consumption goods that are private and public within the household. Following some of the papers that have estimated Euler equations, we will control for potential non-separabilities between consumption and leisure in the estimation. Moreover, since public consumption is an important aspect of household decisions, we will distinguish between private and public goods starting from the next section. In the model, household members live for T periods and can transfer resources over time using a risk-free asset with gross return R. The only source of uncertainty is household income, Y, which is the sum of members' income, $Y = \sum y^i$.

To generalize the collective model to an intertemporal setting, one has to take a stand on the ability of household members to commit to future allocations of resources. We will assume that household members can commit to future plans. The conclusions of this section do not change if a model without commitment is employed.[2] Under full commitment and the assumptions on preferences, the intertemporal collective model can be written in the following form:

$$\max_{\left\{c^1_{t,\omega}, c^2_{t,\omega}, s_t\right\}_{t,\omega}} \quad \mu(Z)\, E_0\left[\sum_{t=0}^{T} \beta^t u^1\left(c^1_{t,\omega}\right)\right] + \left(1-\mu(Z)\right) E_0\left[\sum_{t=0}^{T} \beta^t u^2\left(c^2_{t,\omega}\right)\right] \tag{2.1}$$

$$s.t. \quad \sum_{i=1}^{2} c^i_{t,\omega} + b_t \leq \sum_{i=1}^{2} y^i_{t,\omega} + R_t b_{t-1} \quad \forall t, \omega$$

$$b_T \geq 0 \ \ \forall \omega.$$

The Pareto weights μ and $1-\mu$ can be interpreted as the individual decision power of the two household members. They generally depend variables that have an effect on the intra-household decision power such as prices, wages, and income. We will denote with Z the vector that includes those variables.

[2]The differences between collective models of the household with and without commitment are discussed in Mazzocco (2007) and Chiappori and Mazzocco (2017).

An important feature of the intertemporal collective model just introduced is that its solution is identical to the solution of the following two stage formulation of the household problem. In the first stage, the household allocates optimally lifetime resources across periods and states of nature. In the second stage, conditional on the amount of resources allocated to a given period and state of nature, the household chooses their optimal allocation between the two spouse. To formally describe the two stages, it is convenient to start from the second stage. Denote with $C_{t,\omega}$ an arbitrary amount allocated to period t and state ω. Conditional on $C_{t,\omega}$, in the second stage the household chooses the consumption of the two spouses as the solution of the following static problem:

$$U\left(C_{t,\omega},\mu\left(Z\right)\right)=\max_{\left\{c^1_{t,\omega},c^2_{t,\omega}\right\}}\quad \mu\left(Z\right)u^1\left(c^1_{t,\omega}\right)+\left(1-\mu\left(Z\right)\right)u^2\left(c^2_{t,\omega}\right) \tag{2.2}$$

$$s.t.\quad \sum_{i=1}^{2}c^i_{t,\omega}=C_{t,\omega},$$

where $U\left(C_{t,\omega},\mu\left(Z\right)\right)$ is the household's indirect utility function. In the first stage, the household chooses the allocation of lifetime resources over time and across states of nature using the indirect utility function derived in the second stage by solving the following standard intertemporal problem:

$$\max_{C_{t,\omega}}\quad E_0\left[\sum_{t=0}^{T}\beta^t U\left(C_{t,\omega},\mu\left(Z\right)\right)\right] \tag{2.3}$$

$$s.t.\quad C_{t,\omega}+b_t=Y_{t,\omega}+R_{t+1}b_{t+1}\quad \forall t,\omega. \tag{2.4}$$

The two-stage formulation of the household decision problem is useful because the household Euler equation can easily be derived using the first stage of the household decision process and standard steps. It takes the following form:

$$U_C\left(C_t,\mu\left(Z\right)\right)=\beta R_{t+1}E_0\left[U_C\left(C_{t+1},\mu\left(Z\right)\right)\right]. \tag{2.5}$$

Equation (2.5) can be used to clarify why the model commonly used to derive household Euler equations is problematic if applied to couples. Consider two households in which the two husbands have identical risk aversion, the two wives have identical risk aversion, and the wives' risk aversion is greater than the husbands' risk aversion. The two households are identical in any dimension except that the wife in the first household has a college degree at the time of marriage, whereas the other wife only has a high school diploma. The two households have therefore different vector Z and different intra-household decision power $\mu(Z)$. Suppose that, given the higher education of the wife in the first household, she has more decision power than the wife in the second household. The first household will then be more risk averse than the second one, it will assign more value to consumption smoothing, and it will have a flatter consumption path. The difference in consumption paths will be detected in the estimation of the Euler equations and will be explained using the only variable that can rationalize it, education, and variables that are correlated with it, such as wages and labor force participation. The reduced-form result will therefore be that education, wages, and labor force participation have an effect on the household Euler equations even if they are known at the time decisions are made. The intertemporal model will therefore be rejected because of excess sensitivity.

Notice that the previous argument does not apply to singles, since for them the indirect utility function $U(C_{t,\omega}, \mu(Z))$ simplifies to the standard utility function $U(C_{t,\omega})$. The excess sensitivity test is therefore well-defined for households with only one decision maker. This result provides us with a way of testing our hypothesis that the constant rejection of the household Euler equations is generated by the fact that most households are composed of individuals with different preferences who make joint decisions which depend on their decision power. If this hypothesis is correct, an excess sensitivity test should reject the household Euler equations for the sample of couples, but it should not reject them for the sample of singles. This is the subject of the next two subsections.

2.2.2 The Consumer Expenditure Survey

Since 1980, the CEX survey has been collecting data on household consumption, income, and different types of demographics. The survey is a rotating panel organized by the Bureau of Labor Statistics (BLS). Each quarter about 4500 households representative of the US population are interviewed. 80% are reinterviewed the following quarter, while the remaining 20% are replaced by a new randomly selected group. Each household is interviewed at most for four quarters and detailed information is elicited with regard to expenditures for each of the three months preceding the interview and with regard to income and demographics for the quarter preceding the interview. Information on income is collected only in the first and fourth quarters and it measures income of the year preceding the interview. We assume that the rate at which income is earned is constant for the year. Under this assumption one can construct income for the first and fourth quarters by dividing income by four. The data used in the estimation cover the period 1982-1995. The first two years are excluded because the data were collected with a slightly different methodology.

Following Attanasio and Weber (1995) total consumption is computed as the sum of food at home, food out, tobacco, alcohol, public and private transportation, personal care, maintenance, heating fuel, utilities, housekeeping services, repairs and clothing. As in Attanasio and Weber (1995), the price index is constructed in two steps. First, the components of the Consumer Price Index published by the Bureau of Labor Statistic for each category of consumption are gathered. The price index is then computed as the weighted average of these components where the weights corresponds to the expenditure share spent by a give household on that particular component of consumption. Total consumption is deflated using this household specific price indexes. Household income is computed as total household income plus transfers for the year preceding the interview. The wife's income is the sum of the components that can be imputed to her, i.e. income received from non-farm business, income received from farm business, wage and salary income, social security checks, and supplemental security income checks for the year preceding the interview. As in Attanasio and Weber (1995), the real interest rate is the quarterly average of the 20-year Municipal bond rate deflated using the household specific price index.

Rather than employing the short panel available in the CEX, we follow Attanasio and Weber (1995) and use synthetic panels. These are constructed using two variables: the year of birth of the head of the household and a dummy equal to 1 if the head is married and 0 otherwise.[3] All households are assigned to one of these cells which are constructed using a 7-year interval for the head's year of birth. The variables of interest are then averaged over all the households belonging to a given cohort observed

[3]The husband is assumed to be the head of the household in a married or cohabiting couple.

in a given quarter. To avoid the complicated error structure that the timing of the interviews implies, we follow Attanasio and Weber (1995) and for each household in each quarter we use only the consumption data for the month preceding the interview and drop the data for the previous two months.

To construct the synthetic cohorts we exclude from the sample rural households, households with incomplete income responses, and households experiencing a change in marital status. Only cohorts for which the head's age is between 21 and 60 are included in the estimation. With regard to the minimum cohort size allowed in the estimation, two specifications are employed. In one case, the same size cutoff of 150 observations is used for married and single households. The use of the same cutoff has the disadvantage of generating samples of different size for singles and married. To determine whether the differences in results for married and single households is a consequence of the heterogeneity in sample size, a second specification is used that generates samples with the same number of observations. In this case, for married households, cohorts with size smaller than 150 are dropped. For single households, a cohort is dropped if it has size smaller than 100. An cohort-observation is then kept in a particular period if it is available for both the cohort of singles and the cohort of married individuals. Table 2.1 reports the summary statistics for the CEX sample.

2.2.3 Excess Sensitivity Test for Singles and Couples

In this subsection, we test the formulated hypothesis that the rejections of the Euler equations are generated by the aggregation of individual preferences in households with more than one adult. We do this in two steps. We first follow previous papers and estimate the Euler equations on the sample that includes both singles and couples. We then divide the sample in a subsample composed exclusively of singles and a subsample composed only of couples and estimate the Euler equations separately on the two subsamples. If our hypothesis is correct, we should reject the household Euler equations only for couples. If the rejections are explained by other hypothesis, such as the existence of preference shocks and non-separabilities between consumption and leisure, the Euler equations should be rejected for both couples and singles.

To be consistent with the empirical approach used in the literature on the estimation of Euler equations, we employ the functional form used in most of those papers. Specifically, we estimate Euler equations of the following form:

$$\Delta \log \left(C_{i,t+1} \right) = \alpha + \zeta \log R_{t+1} + \epsilon_{i,t+1}, \tag{2.6}$$

where ζ represents the intertemporal elasticity of substitution, $\epsilon_{i,t+1}$ is a residual which captures the expectational error at time t, and the constant is a function of the discount factor and of the second and higher moments corresponding to the distribution of ϵ_{t+1}. Using standard arguments, this equation can be derived by log-linearizing the Euler equation (2.5) derived earlier in the paper. Here is where our assumption that individuals only have preferences over non-durable consumption, without distinguishing between private and public goods, is important. Absent this assumption, we would not be able to estimate the same Euler equations employed in the literature.

To account for changes in household composition and preference shocks, we follow Attanasio and Weber (1995) and Zeldes (1989) and assume that the demographic variables and preference shocks z enter the instantaneous utility function multiplicatively through an exponential function, which implies that

$$U_{i,t} = U \left(C_{i,t} \right) exp \left(\phi' z_{i,t} \right)$$

In the estimation, the vector z will be composed of family size, number of children, and a set of seasonal dummies.

The Euler equation (2.6) is derived under the assumption that household consumption is strongly separable from leisure. To measure the effect of possible non-separabilities between consumption and leisure, we follow Browning and Meghir (1991), Attanasio and Weber (1995), and Meghir and Weber (1996) and model the leisure variables as conditioning variables, namely variables that may affect preferences over the goods of interest, but are not of primary interest. Specifically, following Attanasio and Weber (1995) and Meghir and Weber (1996), the Euler equations will also be estimated by including different combinations of the following variables: the head's leisure, a dummy equal to 1 if the head works, and, for couples, similar variables for the spouse.

To summarize, the following Euler equation is estimated:

$$\Delta \log \left(C_{i,t+1}\right) = \alpha + \gamma \log R_{t+1} + \phi' \Delta \bar{z}_{i,t+1} + \epsilon_{i,t+1}, \tag{2.7}$$

where $\bar{z}_{i,t+1}$ includes demographic and labor supply variables.

Before reporting the results, we discuss some econometeric issues. The error term of equation (2.7), $\epsilon_{i,t+1}$, contains the expectation error implicit in the Euler equation. Since part of the expectation error is generated by aggregate shocks, $\epsilon_{i,t+1}$ should be correlated across households. This implies that Euler equations can be consistently estimated only if households are observed over a long period of time as suggested by Chamberlain (1984). In the CEX, one of the main advantages of using synthetic panels is that cohorts are followed for the whole sample period. This should reduce the effect of aggregate errors on the estimation results.

The Euler equations are estimated using the Generalized Method of Moments (GMM). Under the assumption of rational expectations, any variable known at time t should be a valid instrument for GMM. However, measurement errors may introduce dependence between variables known at time t and concurrent and future variables even under rational expectations. To address this issue we only use variables known at $t-1$. The GMM estimates are obtained using the efficient weighting matrix, which is generated using a consistent GMM estimator in a first step. Given the longitudinal nature of the dataset employed in the estimation, it is important to allow each household to have a different and unrestricted covariance structure. To that end, the covariance matrix is computed using the efficient weighting matrix in the GMM procedure. As shown in Wooldridge (2002), this covariance matrix is general enough to allow for heteroskedasticity and arbitrary dependence in the residuals.

We can now discuss our empirical findings. The results obtained by estimating the household Euler equations on the sample that includes both singles and couples are reported in Table 2.2. The first three columns describe the results when a two-step GMM estimator is used, whereas in columns 4 through 6 we discuss the results obtained employing a 2-stage least squares estimator to evaluate the effect of the efficient weighting matrix. Since the results are similar, we will only discuss the first set of estimates. In the first column, we report the outcome of the test when we only include demographic variables in the estimation. The results are consistent with the findings of previous papers: the intertemporal elasticity of substitution is positive and around 0.15, but statistically insignificant; changes in family size and number of children have a positive effect on consumption growth; the coefficient on time-t log income is large, positive, and statistically significant, indicating that the excess sensitivity test rejects the intertermporal model. In column 2, we add to the estimation a dummy equal to 1 if the head works and a similar dummy for the spouse as variables capable of capturing potential non-separabilities between consumption and leisure. The coefficient on the head's dummy is negative large and statistically significant, which suggests that households in which the head works are better able to smooth consumption. This variable is therefore capable of explaining part of the variation in consumption growth, but the coefficient on income is even larger than in the first column and statistically significant. In the third column, in addition to the two labor dummies, we add the log of the head's and of spouse's quarterly leisure. The estimated coefficients are consistent with the ones described in column 2. Households in which the head works longer hours are better able to smooth consumption shocks. Remarkably, once we control for the head's leisure, the spouse's work dummy becomes positive and statistically significant. This result can be explained using an added-worker effect argument: with a significant probability, a spouse chooses to participate in the labor market when the households is hit by an adverse shock that increases consumption volatility. In spite of the introduction of the two leisure variables, the coefficient on time t income is still positive, large, and statistically significant, indicating that the household Euler equations are rejected.

We will now describe the results obtained by estimating the household Euler equations separately on the samples of singles and couples, which are reported in Table 2.3. We only present the results obtained using the two-step GMM estimator because the ones obtained using 2-stage least squares are similar. The first column reports the estimates for the sample of singles. The results are similar to the ones obtained for the entire sample except that now the coefficient on income is about two thirds the size of the coefficient estimated for the entire sample and statistically indistinguishable from zero. In this first specification, there is therefore no evidence of excess sensitivity for singles. In the second column, we present the results obtained by estimating the same specialization for couples. The intertemporal elasticity of substitution, the coefficient on family size, and the coefficient on number of children are all estimated to be positive. More importantly, the coefficient on income is positive, more than twice the size the coefficient for singles, and statistically significant. For couples, we can therefore reject this specification of the Euler equations. In columns 3 and 4, we add to the estimation for singles and couples the work dummy for the head of the household and, if present, for the spouse. In both samples, the coefficient on the dummy is positive, large, and statistically significant. This finding can be explained by non-separabilities between consumption and leisure: households with a head that changes employment status from not working to working increase their expenditure on goods related to their job such as gasoline and clothing. The main result, however, does not change. The coefficient on period-t income is still small and insignificant for singles and large and statistically significant for couples. In the last two columns, we report the estimates when we include in the specification the log of the head's quarterly leisure and, for couples, the same variable for the spouse. The leisure variable is positive for the head of a single household, but not statistically significant. For couples, the coefficient on the wife's leisure is positive and large, but statistically insignificant. Similarly to the previous specification, the coefficient on income is negligible and insignificant for singles, but large, positive, and statistically significant for couples.

We can therefore conclude that, even with the addition of demographic variables and variables that account for possible non-separabilities between consumption and leisure, Euler equations are rejected for couples because of excess sensitivity. We cannot, however, reject them for singles. These findings indicate that the excess sensitivity displayed by the Euler equations when estimated on the entire sample is generated by couples. As such, they provide strong evidence in favor of our hypothesis that the rejection of the Euler equations is produced by the aggregation of individual preferences in households with several decision makers.

Recently, a growing number of papers in micro and macro economics have developed intertermporal models of the households in which household members are characterized using individual utility functions.[4] To answer relevant questions, these papers require reliable estimates of the intertemporal preferences for women and men. The results of this section imply that the parameters characterizing the intertemporal preferences, such as the intertemporal elasticity of substitution, cannot be estimated using household

[4]Some examples are Casanova (2010), Gemici and Laufer (2012), Voena (2015), Greenwood et al. (2003), and Doepke and Tertilt (2019).

Euler equations and samples composed of all households or samples composed of only couples. One possible way of addressing this issue is to rely only on singles for which households Euler equations are not rejected. One limitation of this approach is that it only works if there is no selection into marriage based on intertemporal preferences. For instance, if the more risk averse women and the less risk averse men are less likely to find a partner, the intertemporal preferences estimated on the sample of singles are not well suited to characterize the intertemporal preferences of married individuals. In the rest of the paper, we develop an alternative method to identify and estimate the intertemporal preferences of married women and men which is not affected by the selection issue mentioned above. Our method will also enable us to understand whether selection into marriage is a real concern when estimating intertemporal preferences. If it is not, the sample of singles can safely be used to recover the intertemporal aspects of individual preferences.

2.3 Household and Individual Euler Equations

Consider a two-person household living for $\mathcal{T}$ periods in an uncertain environment. In each period $t \in \{0, ..., \mathcal{T}\}$ and state of nature $\omega \in \Omega$, member i receives non-labor income $y^i(t,\omega)$, supplies labor in quantity $h^i(t,\omega)$, and chooses expenditure on a private composite good $c^i(t,\omega)$ and on children $Q(t,\omega)$. Since children are for the most part a public good for their parents, $Q(t,\omega)$ will be modeled as public consumption. Let $C(t,\omega)$ be household total private consumption and let $l^i(t,\omega) = 1 - h^i(t,\omega)$ be leisure of member i, where the time endowment is normalized to 1. The price of private and public consumption will be denoted by $p(t,\omega)$ and $P(t,\omega)$, and agent i's wage by $w_i(t,\omega)$. Household members can save jointly using a risk-free asset. Denote by $s(t,\omega)$ and $R(t)$, respectively, the amount of wealth invested in the risk-free asset and its gross return.[5] Each household member is characterized by individual preferences, which are assumed to be separable over time and across states of nature. The corresponding utility function U^i is assumed to be increasing, concave, and twice continuously differentiable. The corresponding utility function U_i is assumed to be increasing, concave, and twice continuously differentiable. Agent i's utility function can depend on agent j's private

[5]The results of the paper are still valid if risky assets are introduced in the model.

consumption and leisure but only additively, i.e.

$$U^{i}\left(c^{1},c^{2},l^{1},l^{2},Q\right)=u^{i}\left(c^{i},l^{i},Q\right)+\delta_{i}u^{j}\left(c^{j},l^{j},Q\right),$$

where δ_i is the altruism parameter. It is assumed that the two spouses have the same discount factor β.[6]

The next two subsections describe two different approaches to identifying and estimating the intertemporal and intratemporal features of the preferences that characterize household decisions.

2.3.1 Household Euler Equations

The theoretical and empirical literature on intertemporal decisions has traditionally assumed that households behave as single agents independently of the number of decision makers. This is equivalent to assuming that the utility functions of the individual members can be collapsed into a unique utility function which fully describes the preferences of the entire household. Following this approach, suppose that household preferences can be represented using a unique von Neumann-Morgenstern utility function $U\left(C,l^{1},l^{2},Q\right)$ and a household discount factor β. Intertemporal decisions can then be determined by solving the following problem:[7]

$$\max_{\left\{C_t,l_t^1,l_t^2,Q_t,s_t\right\}} E_0\left[\sum_{t=0}^{\mathcal{T}}\beta^{t}U\left(C_t,l_t^1,l_t^2,Q_t\right)\right] \tag{2.8}$$

$$s.t.\ p_tC_t+P_tQ_t+s_t\leq\sum_{i=1}^{2}\left(y_t^i+w_t^ih_t^i\right)+R_ts_{t-1}\quad\forall t,\omega$$

$$s_T\geq 0\quad\forall\omega.$$

[6]This assumption is made for expositional purposes. If the individual discount factors are different, it can be shown that the identification method proposed here still works with small modifications.

[7]The dependence on the states of nature is suppressed to simplify the notation.

The first order conditions of the unitary model (2.8) can be used to derive the following standard household Euler equations for private consumption:

$$U_C\left(C_t, l_t^1, l_t^2, Q_t\right) = \beta E_t\left[U_C\left(C_{t+1}, l_{t+1}^1, l_{t+1}^2, Q_{t+1}\right) R_{t+1}\frac{p_t}{p_{t+1}}\right].$$

Since the variables defining these intertemporal optimality conditions are observed in various datasets, in the past two decades the standard household Euler equations have been used to test the intertemporal decisions of the household and to estimate the parameters that characterize its behavior.

This approach has one major limitation: the parameter estimates of the intertemporal unitary model can be used to understand household behavior and to answer policy questions only if households behave as single agents. Mazzocco (2007) shows that this assumption is satisfied if and only if the following strong restrictions on individual preferences are satisfied: (i) household members have identical discount factors; (ii) the individual preferences belong to the HARA class and have identical curvature parameters. The evidence based on household Euler equations indicates that this assumption is violated. In particular, in the past twenty years economists have rejected household Euler equations using either the sample of couples or the sample of couples jointly with singles.[8] Two additional tests based on household Euler equations are performed in Mazzocco (2007) and the outcome suggests that the behavior of a group of agents differs from the behavior of single agents. Additional evidence against the unitary model has been collected in a static framework for instance by Thomas (1990), Browning, Bourguignon, Chiappori and Lechene (1994), Browning and Chiappori (1998).

These empirical findings indicate that it may be important to estimate an alternative model that better characterizes the intertemporal behavior of the household.

2.3.2 Individual Euler Equations

This section relaxes the assumption that the individual utility functions can be collapsed into a unique utility function. Without this restriction, it must be established

[8]See Browning and Lusardi (1996) for a survey.

how individual preferences are aggregated to determine household decisions. Following Chiappori (1988; 1992) and Mazzocco (2004; 2007), it is assumed that every decision is on the ex-ante Pareto frontier, which implies that household intertemporal behavior can be characterized as the solution of the following Pareto problem:

$$\max_{\{c_t^1, c_t^2, l_t^1, l_t^2, Q_t, s_t\}} \mu E_0 \left[\sum_{t=0}^{\mathcal{T}} \beta^t u^1(c_t^1, l_t^1, Q_t) \right] + (1-\mu) E_0 \left[\sum_{t=0}^{\mathcal{T}} \beta^t u^2(c_t^2, l_t^2, Q_t) \right] \tag{2.9}$$

$$s.t. \sum_{i=1}^{2} \left(p_t c_t^i + w_{it} l_t^i \right) + P_t Q_t + s_t = \sum_{i=1}^{2} \left(y_t^i + w_{it} \right) + R_t s_{t-1} \ \ \forall t, \omega$$

$$0 \leq l_t^i \leq 1 \ \ \forall i, t, \omega, \qquad s_T \geq 0 \ \ \forall \omega,$$

where μ is a combination of Pareto weights and altruism parameters, and it can be interpreted as the relative decision power at the time of household formation.

Under standard assumptions, the following Euler equations for consumption can be derived:

$$u_c^i \left(c_t^i, l_t^i, Q_t \right) = \beta_i E_t \left[u_c^i \left(c_{t+1}^i, l_{t+1}^i, Q_{t+1} \right) R_{t+1} \frac{p_t}{p_{t+1}} \right] \ \ \forall \ i = 1, 2. \tag{2.10}$$

Two remarks are in order. First, the individual Euler equations are not affected by the aggregation problem that affects the standard household Euler equations, since they are satisfied independently of the number of household members. Second, the leisure Euler equations could be added to the consumption Euler equations to characterize the intertemporal behavior of the household. However, they are satisfied only if the corresponding agent supplies a positive amount of labor in each period and state of nature. Since this assumption is excessively strong, only the consumption Euler equations will be employed.

To discuss the identification of individual preferences it is helpful to rewrite the household problem using a two-stage formulation. Under the assumption that individual preferences are separable over time and across states of nature, the solution of the household problem (2.9) is equivalent to the solution of the following two-stage problem. In the second stage, conditional on the amount of resources available in period t and state ω, the household chooses how much to spend on consumption and leisure.

Formally, let $\bar{Y}(t, \omega)$ be the amount of resources available in period t and state ω. In the second stage, the household solves the following static problem for each t and ω:

$$
\begin{aligned}
V\left(\bar{Y}_t, w_{1t}, w_{2t}, p_t, P_t\right) = & \max_{c_t^1, l_t^1, c_t^2, l_t^2, Q_t} \mu u^1\left(c_t^1, l_t^1, Q_t\right) + (1-\mu) u^2\left(c_t^2, l_t^2, Q_t\right) \\
& s.t. \sum_{i=1}^{2}\left(p_t c_t^i + w_{it} l_t^i\right) + P_t Q_t = \bar{Y}_t \\
& 0 \leq l_t^i \leq 1 \quad \text{for } i = 1, 2.
\end{aligned}
$$

The standard Marshallian demand functions for public consumption, household private consumption, and leisure can be derived as the solution of this second-stage problem. They depend on the prices of public and private consumption, the individual wages, and the resources available in period t and state ω, i.e. $Q_t = Q\left(p_t, P_t, w_{1t}, w_{2t}, \bar{Y}_t\right)$, $C_t = c_t^1 + c_t^2 = C\left(p_t, P_t, w_{1t}, w_{2t}, \bar{Y}\right)$, $l_t^1 = l^1\left(p_t, P_t, w_{1t}, w_{2t}, \bar{Y}_t\right)$, and $l_t^2 = l^2\left(p_t, P_t, w_{1t}, w_{2t}, \bar{Y}_t\right)$. In the first stage the household chooses the optimal allocation of resources to each period and state of nature by solving the following dynamic problem:

$$
\begin{aligned}
& \max_{\left\{\bar{Y}_t, s_t\right\}} \sum_{t=0}^{\mathcal{T}} E_0\left[\beta^t V\left(\bar{Y}_t, w_{1t}, w_{2t}, p_t, P_t\right)\right] \\
& s.t.\ \bar{Y}_t = \sum_{i=1}^{2}\left(y_t^i + w_{it}\right) + R_t s_{t-1} - s_t \quad \forall t, \omega \\
& s_T \geq 0 \ \forall \omega.
\end{aligned}
$$

The two-stage formulation will be used to describe the type of variation required in the identification of the individual preferences for expenditure on children.

Three main assumptions characterize the intertemporal collective model (2.9) and the corresponding Euler equations. First, the household Euler equations as well as the individual Euler equations characterize only the intertemporal behavior of households that are not borrowing constrained. There is mixed evidence on the importance of liquidity constraints. For instance, Zeldes (1989) and Gross and Souleles (2002) find that borrowing constraints characterize a significant fraction of the U.S. population. Meghir and Weber (1996) find that at most a small fraction of households are liquidity constrained. The theoretical and empirical results of this paper hold for household that are not borrowing constrained in the period considered in the analysis.[9]

Second, it is assumed that there is no household production or equivalently that household production is determined exogenously. Under this assumption, if individual labor supply is observed, individual leisure is also observed. The generalization of the identification and estimation results to a framework with household production is an important research topic, but it is left for future research.

Third, it is assumed that household decisions are always on the ex-ante Pareto frontier, which implies that the individual members must be able to commit to future allocations of resources at the time of household formation. To test whether the assumption of ex-ante efficiency represents a good approximation of household decisions the following standard efficiency condition will be analyzed jointly with the Euler equations:

$$\frac{u_c^1\left(c_t^1, l_t^1, Q_t\right)}{u_c^2\left(c_t^2, l_t^2, Q_t\right)} = \mu. \tag{2.11}$$

If individual private consumption and individual labor supply were observed, individual preferences could be estimated using the individual Euler equations and the efficiency condition. Unfortunately, consumption is only measured at the household level. The next section is devoted to showing that the parameters that characterize household intertemporal behavior can be identified using the consumption Euler equations and the intraperiod conditions even if consumption is not observed at the individual level.

2.4 A General Example

In this section we will consider an example which illustrates how the intertemporal conditions help in the identification of the parents' preferences for expenditure on children. In an attempt to provide a clear intuition of the identification results, in this session we

[9]Future borrowing constraints affect household decisions in period t and $t+1$. This effect is captured in the individual Euler equations by the information set at t. Consequently, as long as the individual Euler equations are satisfied during the survey period, the identification and estimation results hold.

will consider an environment with no uncertainty. we will only discuss the case in which only agent 1, the father, works since in this case the identification of the parameters of interest is more difficult to achieve.

In the example discussed in this section we will consider a specific functional form for the parents' preferences. It is assumed that each parent has a utility function that is non-separable in public consumption and has the following form:

$$U\left(l^i, c^i, Q\right) = \left(c^i\right)^{\sigma_i}\left(l^i\right)^{\theta_i}(Q)^{\gamma_i} + \delta_i \ln Q.$$

The utility function is a standard Cobb-Douglas utility function augmented to include a public good. The public good enters individual preferences in two different ways: through a separable function and through a function in which public consumption is non-separable from leisure and private consumption. This feature of the utility function will enable me to describe which variation is required for the identification of the non-separable part of the preferences for public consumption and which variation is needed to recover the separable part.

The problem of identifying the parameters of interest can be stated in the following way. The econometrician knows public consumption, household private consumption, the father's leisure, the father's wage, the prices of private and public consumption, and the amount of resources that the household decides to allocate to each period. Since the mother does not work, no variation in her wage and leisure is observed. A dataset in which all these variables are observed is the Consumer Expenditure Survey (CEX). Using these variables, the econometrician can recover non-parametrically the Marshallian demand functions $Q = Q\left(p, P, w_1, \bar{Y}\right)$, $C = C\left(p, P, w_1, \bar{Y}\right)$, and $l^1 = l^1\left(p, P, w_1, \bar{Y}\right)$, which are the solution of the second stage of the household problem. They will therefore be assumed to be known. Note that the Marshallian demand function for the mother's leisure cannot be recovered because by assumption there is no variation in her leisure. Since the Marshallian demand functions are known, the derivatives of public consumption, private consumption, and the father's leisure with respect to wages, resources, and prices are also known. Given this information, the econometrician is interested in recovering the preference parameters and the decision power parameter.

In the parametric examples considered in this section, identification can be easily analyzed using the first order conditions for private consumption, leisure, and public consumption. The following approach will be employed. The first order conditions will be used to derive a set of equations that depend on the parameters of interest and variables that are known. The equations can then be solved for the parameters of interest. If a unique solution exists, the model is identified. If more than one solution exist, the model is not identified.

We will start with the derivation of the first order conditions. Denote by λ_t the multiplier of the budget constraint of the household problem in period t and let $\mu_1 = \mu$ and $\mu_2 = 1 - \mu$. In the example considered here, the first order conditions for private consumption of parent i can be written in the form

$$\beta^t \mu_i \sigma_i \left(c_t^i\right)^{\sigma_i - 1} \left(l_t^i\right)^{\theta_i} \left(Q_t\right)^{\gamma_i} = p_t \lambda_t,$$

the leisure first order condition of the working parent takes the form

$$\beta^t \mu_1 \theta_1 \left(c_t^1\right)^{\sigma_1} \left(l_t^1\right)^{\theta_1 - 1} \left(Q_t\right)^{\gamma_1} = w_{1t} \lambda_t,$$

and the public consumption first order condition can be written as

$$\beta^t \sum_{i=1}^{2} \mu_i \left(\gamma_i \left(c_t^i\right)^{\sigma_i} \left(l_t^i\right)^{\theta_i} \left(Q_t\right)^{\gamma_i - 1} + \frac{\delta_i}{Q_t} \right) = P_t \lambda_t,$$

where $l_t^2 = 1$ for the mother. Finally, in an environment without uncertainty, the first order condition that captures the optimal allocation of resources over time has the following form:

$$\lambda_t = R_{t+1} \lambda_{t+1}.$$

Using these first order conditions one can derive the five optimality conditions that will be employed in the identification of the parameters of interest: (i) an equation stating that the marginal rate of substitution between consumption and leisure of the working parent must equal his real wage; (ii) the efficiency condition for private consumption; (iii) the efficiency condition for public consumption; (iv) the private consumption Euler equation for the mother; (v) the private consumption Euler equation for the father.[10] In the present example, the first optimality condition has the following form:

$$\frac{\theta_1}{\sigma_1}\frac{c_t^1}{l_t^1} = \frac{w_{1t}}{p_t}.$$

The private consumption efficiency condition can be written as

$$\frac{\sigma_1\,(c_t^1)^{\sigma_1-1}\,(l_t^1)^{\theta_1}\,Q_t^{\gamma_1}}{\sigma_2\,(c_t^2)^{\sigma_2-1}\,Q_t^{\gamma_2}} = \frac{1-\mu}{\mu}. \tag{2.12}$$

Using the first order conditions for public and private consumption, the public consumption efficiency condition can be written as follows:

$$\frac{\gamma_1}{\sigma_1}\frac{c_t^1}{Q_t} + \frac{\delta_1}{\sigma_1}\frac{1}{(c_t^1)^{\sigma_1-1}\,(l_t^1)^{\theta_1}\,Q_t^{\gamma_1+1}} + \frac{\gamma_2}{\sigma_2}\frac{c_t^2}{Q_t} + \frac{\delta_2}{\sigma_2}\frac{1}{(c_t^2)^{\sigma_2-1}\,Q_t^{\gamma_2+1}} = \frac{P_t}{p_t}.$$

Finally, the father's private consumption Euler equation takes the form

$$\beta R_{t+1}\frac{p_t}{p_{t+1}}\left(\frac{c_{t+1}^1}{c_t^1}\right)^{\sigma_1-1}\left(\frac{l_{t+1}^1}{l_t^1}\right)^{\theta_1}\left(\frac{Q_{t+1}}{Q_t}\right)^{\gamma_1} = 1,$$

whereas the mother's can be written as follows:

$$\beta R_{t+1}\frac{p_t}{p_{t+1}}\left(\frac{c_{t+1}^2}{c_t^2}\right)^{\sigma_2-1}\left(\frac{Q_{t+1}}{Q_t}\right)^{\gamma_2} = 1.$$

We will now discuss how the parameters of the non-separable part of the father's preferences σ_1, θ_1, and γ_1 can be recovered by using his private consumption Euler equation. The Euler equations depend on private consumption which is not observed. However, one can use the optimality condition that relates the marginal rate of substitution of the father to his real wage to derive the father's private consumption as a

[10]The model considered in this paper is over-identified in the sense that the number of optimality conditions that can be used to recover the parameters of interest is greater than the number of parameters. For instance, the public consumption Euler equation could be used in place of one of the five conditions employed in this section. The optimality conditions that are not used in the identification of the parameters can be used to test the model.

function of own leisure and own real wage, i.e.

$$c_t^1 = \frac{\sigma_1}{\theta_1}\frac{w_{1t}}{p_t}l_t^1.$$

In each period the sum of individual private consumption must equal total household private consumption C_t, which is observed. As a consequence, the mother's private consumption can also be written as a function of variables that are observed, i.e.

$$c_t^2 = C_t - \frac{\sigma_1}{\theta_1}\frac{w_{1t}}{p_t}l_t^1.$$

The private consumption Euler equation of the father can now be written in terms of variables that are observed by substituting out individual consumption. After the substitution, this intertemporal optimality condition becomes

$$\beta R_{t+1}\left(\frac{p_t}{p_{t+1}}\right)^{\sigma_1}\left(\frac{w_{1t+1}}{w_{1t}}\right)^{\sigma_1-1}\left(\frac{l_{t+1}^1}{l_t^1}\right)^{\sigma_1+\theta_1-1}\left(\frac{Q_{t+1}}{Q_t}\right)^{\gamma_1} = 1.$$

By taking the logarithm of both sides, it can be rewritten in the following simpler form:

$$\Delta \ln Q_{t+1} + \rho_1^1 \Delta \ln l_{t+1}^1 = -\rho_2^1 \Delta \ln w_{1t+1} - \rho_3^1 \ln\left(\frac{p_t}{p_{t+1}}\right) - \rho_4^1 \ln\left(R_{t+1}\right) - \rho_4^1 \ln \beta. \quad (2.13)$$

where $\rho_1^1 = \frac{\sigma_1+\theta_1-1}{\gamma_1}$, $\rho_2^1 = \frac{\sigma_1-1}{\gamma_1}$, $\rho_3^1 = \frac{\sigma_1}{\gamma_1}$, and $\rho_4^1 = \frac{1}{\gamma_1}$. Two features of this optimality condition are worth a discussion. First, this equation depends on five unknown parameters ρ_1^1, ρ_2^1, ρ_3^1, ρ_4^1, and β, and observed variables. Second, in an environment with uncertainty one needs this equation plus four additional moment conditions to identify the five parameters. The standard approach in the estimation of parameters contained in Euler equations is to use lagged variables to construct the four additional moment conditions. With the goal of providing some insight on the variation required in the data to identify these parameters, instead of considering the standard case with uncertainty we will discuss the case of no uncertainty.

Consider a change in the amount of resources allocated to period t, $\bar{Y}_t$, generated for instance by a variation in the father's wage in period $t' \neq t,\ t+1$. The household will respond to this variation by changing how the father's leisure, the father's private consumption, and public consumption evolve between t and $t+1$. This intertemporal change depends on the father's taste for leisure and private consumption relative to his taste for public consumption, which is described by ρ_1^1. It can be determined by differentiating the father's private consumption Euler equation (2.13) with respect to $\bar{Y}_t$ and it can be described using the following equation:

$$\Delta \ln Q_{\bar{Y}_t,t+1} + \rho_1^1 \Delta \ln l_{\bar{Y}_t,t+1}^1 = 0,$$

where $Q_{\bar{Y}_t,t+1}$ and $l_{\bar{Y}_t,t+1}^i$ are the partial derivatives of public consumption and leisure with respect to $\bar{Y}_t$. This implies that one can recover the father's taste for leisure and private consumption relative to his taste for public consumption by simply observing the intertemporal change in father's leisure and public consumption in response to a change in resources available in a given period, i.e.

$$\rho_1^1 = -\frac{\Delta \ln Q_{\bar{Y}_t,t+1}}{\Delta \ln l_{\bar{Y}_t,t+1}^1}.$$

Now that the relative taste parameter ρ_1^1 is known, it is straightforward to identify the parameter ρ_3^1 which provides information on the father's taste for private consumption relative to his taste for public consumption. Consider a change in the price of private consumption at t. The household varies the father's leisure and public consumption according to the following optimality condition:

$$\Delta \ln Q_{p_t,t+1} + \rho_1^1 \Delta \ln l_{p_t,t+1}^1 = -\rho_3^1 \frac{1}{p_t}.$$

As a consequence, ρ_3^1 can be recovered if one observes a change in p_t and the corresponding change in leisure and public consumption. Specifically,

$$\rho_3^1 = -p_t \left(\Delta \ln Q_{p_t,t+1} + \rho_1^1 \Delta \ln l_{p_t,t+1}^1 \right).$$

Finally one can recover the parameter ρ_2^1, which provides different information on the father's taste for private consumption relative to his taste for public consumption, if variation in the father's wage in period t is observed. The effect of this variation on the father's leisure and public consumption can be determined by differentiating the father's private consumption Euler equation with respect to his wage. The following equation describes the effect:

$$\Delta \ln Q_{w_{1t},t+1} + \rho_1^1 \Delta \ln l^1_{w_{1t},t+1} = \rho_2^1 \frac{1}{w_{1t}}.$$

The parameter ρ_2^1 is therefore equal to

$$\rho_2^1 = w_{1t} \left(\Delta \ln Q_{w_{1t},t+1} + \rho_1^1 \Delta \ln l^1_{w_{1t},t+1} \right).$$

Now that the reduced form parameters ρ_1^1, ρ_2^1, and ρ_3^1 are known, it is straightforward to recover the father's preference parameters for the non-separable part of his utility function. They are equal to the following functions of the reduced form parameters:

$$\sigma_1 = \frac{\rho_3^1}{\rho_3^1 - \rho_2^1}, \quad \theta_1 = \frac{\rho_1^1 - \rho_2^1}{\rho_3^1 - \rho_2^1}, \quad \gamma_1 = \frac{1}{\rho_3^1 - \rho_2^1}.$$

Since γ_1 has been recovered, the reduced-form parameter ρ_4^1 is also known. The discount factor can then be identified by solving the private consumption Euler equation for β.

The father's private consumption can also be recovered since it only depends on the parameters σ_1 and θ_1. As a consequence, the mother's private consumption is also identified. It should be remarked that individual consumption can be identified only because of the particular functional form chosen for the utility functions. In general, individual consumption can be identified only up to an additive constant. In the example considered here, the constant is assumed to be zero.

We will now describe how the mother's preference parameters for private consumption and the non-separable part of her preferences for public consumption can be recovered using her private consumption Euler equation. Her taste for leisure, however, cannot be identified since no variation in her labor supply is observed. Since the mother's private consumption is now known, there is no need to substitute out private consumption from her Euler equation, which can be written in the form

$$\Delta \ln Q_{t+1} + \rho_1^2 \Delta \ln c^2_{t+1} = -\rho_2^2 \ln \left(R_{t+1} \frac{p_t}{p_{t+1}} \right) - \rho_2^2 \ln \beta,$$

where $\rho_1^2 = \frac{\sigma_2 - 1}{\gamma_2}$ and $\rho_2^2 = \frac{1}{\gamma_2}$. The identification of the mother's preference parameters can be achieved using the logic used for the father. Consider a change in the resources available in period t generated by a variation in one of the exogenous variables in period $t' \neq t,\ t+1$. This change modifies how the household allocates resources between t and $t+1$. The corresponding intertemporal change for the mother can be described by differentiating her private consumption Euler equation with respect to $\bar{Y}_t$, i.e.

$$\Delta \ln Q_{\bar{Y}_t,t+1} + \rho_1^2 \Delta \ln c^2_{\bar{Y}_t,t+1} = 0.$$

This type of variation enables one to recover the mother's taste for private consumption relative to her taste for public consumption ρ_1^2, i.e.

$$\rho_1^2 = -\frac{\Delta \ln Q_{\bar{Y}_t,t+1}}{\Delta \ln c^2_{\bar{Y}_t,t+1}},$$

The inverse of the taste for public consumption ρ_2^2 can now be recovered if variation in the price of private consumption at t and the corresponding changes in intertemporal decisions are observed. These changes are described by the following equation:

$$\Delta \ln Q_{p_t,t+1} + \rho_1^2 \Delta \ln c^2_{p_t,t+1} = -\rho_2^2 \frac{1}{p_t}.$$

which implies that

$$\rho_2^2 = -p_t \left(\Delta \ln Q_{p_t,t+1} + \rho_1^2 \Delta \ln c^2_{p_t,t+1} \right).$$

Finally, the mother's preference parameters for private and public consumption can be recovered using the information on the reduced-form parameters ρ_1^2 and ρ_2^2. Specifically,

$$\sigma_2 = \frac{\rho_1^2 + \rho_2^2}{\rho_2^2}, \quad \gamma_2 = \frac{1}{\rho_2^2}.$$

Only three of the parameters of interest remain to be identified: the decision power parameter μ and the parameters that describe the separable part of the individual preferences for public consumption δ_1 and δ_2. Intuitively, one should expect that these parameters cannot be identified by simply using the private consumption Euler equations, since they provide no information on the individual decision power and on the separable part of the preferences for the public good. Some additional restrictions imposed by the model on individual behavior must be employed to identify the remaining parameters.

The decision power parameter can be recovered using the private consumption efficiency condition (2.12). In this equation all the parameters and variables are known except μ. One can therefore identify the individual decision power by solving this equation for μ. The parameters δ_1 and δ_2 can be recovered using the public consumption efficiency condition in two different periods. The public consumption efficiency condition at t and $t+1$ can be written in the following form:

$$A_t^1 + \delta_1 B_t^1 + A_t^2 + \delta_2 B_t^2 = \frac{P_t}{p_t},$$

and

$$A_{t+1}^1 + \delta_1 B_{t+1}^1 + A_{t+1}^2 + \delta_2 B_{t+1}^2 = \frac{P_{t+1}}{p_{t+1}}.$$

where A_t^i, B_t^i, A_{t+1}^i and B_{t+1}^i are functions of known parameters and observed variables. The parameters that characterize the separable part of the individual preferences for public consumption can therefore be recovered by solving these two equations for δ_1 and δ_2. They can be written in the form

$$\delta_1 = \frac{B_t^2 A_{t+1}^1 + B_t^2 A_{t+1}^2 - A_t^1 B_{t+1}^2 - A_t^2 B_{t+1}^2 - B_t^2 P_{t+1} + P_t B_{t+1}^2}{B_t^1 B_{t+1}^2 - B_{t+1}^1 B_t^2},$$

and

$$\delta_2 = \frac{B_t^1 A_{t+1}^1 + B_t^1 A_{t+1}^2 - B_{t+1}^1 A_t^1 - B_{t+1}^1 A_t^2 + B_{t+1}^1 P_t - B_t^1 P_{t+1}}{B_{t+1}^1 B_t^2 - B_t^1 B_{t+1}^2}.$$

The results presented in this section suggest that the parents' preferences for expenditure on children can be identified even if only one parent works and the individual preferences are non-separable in public consumption. The information on preferences for expenditure on children can be used to predict how much parents in different income quartiles will invest in their children. Policies that attempt to improve the cognitive and non-cognitive abilities of children living in deprived environments can then be designed to reflect potential differences in early investments across income quartiles.

2.5 A General Identification Result

The identification result presented in the previous section will be extended to a general set of utility functions. Identification is achieved in four steps. In the first step, individual consumption is derived as a function of observed variables using the optimality condition that relates the individual marginal rate of substitution between consumption and leisure to own real wage. In the second step, individual consumption is substituted out of the individual marginal utilities using the consumption function obtained in the first step. In the third step, intra-period and intertemporal optimality conditions are derived using the reduced-form marginal utilities obtained in the second step. It is then shown that the reduced-form marginal utilities and individual decision power can be identified using this set of conditions. In the last step, the individual utilities are recovered exploiting the information on the reduced-form marginal utility functions.

Suppose that in each period at least one agent chooses to supply a positive amount of labor. Without loss of generality, it will be assumed that agent 1 satisfies this restriction. Under this assumption, the first order conditions at t for the intertemporal collective model imply that agent 1's marginal rate of substitution between private consumption and leisure must equal the real wage, i.e.,

$$\frac{u_l^1\left(c_t^1, 1-h_t^1, Q_t\right)}{u_c^1\left(c_t^1, 1-h_t^1, Q_t\right)} = q\left(c_t^1, h_t^1, Q_t\right) = \bar{w}_{1t},$$

where $\bar{w}_{1t} = \dfrac{w_{1t}}{p_t}$. If the inverse function of q is well-defined, agent 1's consumption can be written as the following unknown function of individual labor supply, public consumption, and real wage:[11]

$$c_t^1 = g\left(\bar{w}_{1t}, h_t^1, Q_t\right).$$

[11] The consumption function g is well-defined if the marginal rate of substitution q is strictly increasing in consumption, which is a standard assumption in the labor literature. More formally, lemma 1 in the appendix shows that g is well-defined if

$$u_{lc}^1\left(c_t^1, 1-h_t^1, Q_t\right) u_c^1\left(c_t^1, 1-h_t^1, Q_t\right) - u_{cc}^1\left(c_t^1, 1-h_t^1, Q_t\right) u_l^1\left(c_t^1, 1-h_t^1, Q_t\right) \neq 0. \tag{2.14}$$

The function g corresponds to the m-consumption function introduced by Browning (1998).

Since household private consumption is observed and in each period $C_t = c_t^1 + c_t^2$, agent 2's private consumption can also be written as a function of observed variables as follows:

$$c_t^2 = C_t - g\left(\bar{w}_{1t}, h_t^1, Q_t\right).$$

Using the function g, the unobserved individual private consumption can be substituted out of the marginal utilities that define the intratemporal and intertemporal optimality conditions. Denote with f_k^1 and f_k^2 the reduced-form marginal utilities with respect to good k for agent 1 and 2 obtained with this substitution. Then f_k^1 and f_k^2 can be defined as follows:

$$f_k^1\left(\bar{w}_{1t}, h_t^1, Q_t\right) = u_k^1\left(g\left(\bar{w}_{1t}, h_t^1, Q_t\right), 1 - h_t^1, Q_t\right) \quad k = c, l, Q, \tag{2.15}$$

$$f_k^2\left(C_t, \bar{w}_{1t}, h_t^1, h_t^2, Q_t\right) = u_k^2\left(C_t - g\left(\bar{w}_{1t}, h_t^1, Q_t\right), 1 - h_t^2, Q_t\right) \quad k = c, l, Q. \tag{2.16}$$

An example for f_k^1 and f_k^2 can be easily derived using the parametric specification assumed in the previous section. For instance, in that case the father's reduced-form marginal utility for private consumption is characterized by the reduced form parameters $\alpha_1 = \sigma_1 - 1$, $\alpha_2 = \sigma_1 + \theta_1 - 1$, $\alpha_3 = \sigma_1 \left(\frac{\sigma_1}{\theta_1}\right)^{\alpha_1}$, and by the preference parameter γ_1.

The reduced-form marginal utilities can be used to rewrite the individual private consumption Euler equations in terms of observed variables. To that end, the assumption that agent 1 supplies a positive amount of labor must be fulfilled for two consecutive periods. Under this restriction, the intertemporal optimality conditions can be written as follows:

$$f_c^1\left(\bar{w}_{1t}, h_t^1, Q_t\right) = \beta E_t\left[f_c^1\left(\bar{w}_{1t+1}, h_{t+1}^1, Q_t\right) R_{t+1}\frac{p_t}{p_{t+1}}\right],$$

$$f_c^2\left(C_t, \bar{w}_{1t}, h_t^1, h_t^2, Q_t\right) = \beta E_t\left[f_c^2\left(C_{t+1}, \bar{w}_{1t+1}, h_{t+1}^1, h_{t+1}^2, Q_t\right) R_{t+1}\frac{p_t}{p_{t+1}}\right].$$

Since household private consumption, public consumption, individual labor supply, individual wages, and the interest rate are observed, the reduced-form marginal utilities f_c^1 and f_c^2, and the discount factor β can be identified using the private consumption Euler equations and methods that have been developed for the identification of Euler equations.[12]

The remaining reduced-form marginal utilities can be identified using the intra-period optimality conditions and the public consumption Euler equation. Observe that agent 1's marginal rate of substitution between private consumption and leisure must be equal to the real wage even if individual consumption is substituted out using the consumption function g. This implies that

$$\frac{u_l^1\left(g\left(\bar{w}_{1t}, h_t^1, Q_t\right), 1-h_t^1, Q_t\right)}{u_c^1\left(g\left(\bar{w}_{1t}, h_t^1, Q_t\right), 1-h_t^1, Q_t\right)} = \frac{f_l^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)}{f_c^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)} = \bar{w}_{1t}.$$

Now consider a realization of the exogenous variables p_t, P_t, w_{1t}, w_{2t}, and of the amount of resources $\bar{Y}_t$, which are all observed.[13] Conditional on this realization, the household members choose the optimal amount of C_t, Q_t, h_t^1, and h_t^2, which are also observed. For the observed $\bar{w}_{1t}$, h_t^1, Q_t, the function $f_c^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)$ is known from the private consumption Euler equations. Consequently, one can recover $f_l^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)$ for the observed $\bar{w}_{1t}$, h_t^1, and Q_t by setting it equal to $\bar{w}_{1t} f_c^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)$. By using the same argument for every realization of p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$, the entire function f_l^1 can be identified.

The individual decision power can be identified using a similar idea. The private consumption efficiency condition can be written using the reduced-form marginal utilities in the following form:

$$\frac{f_c^1\left(\bar{w}_{1t}, h_t^1, Q_t\right)}{f_c^2\left(C_t, \bar{w}_{1t}, h_t^1, h_t^2, Q_t\right)} = \frac{1-\mu}{\mu}.$$

For every realization of p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$, the functions f_c^1 and f_c^2 are known from the Euler equations. The relative decision power μ is therefore also identified.

The reduced-form marginal utilities of public consumption can be recovered using

[12] As mentioned in the introduction, in principle the Euler equations can be identified non-parametrically. However, until a paper on non-parametric identification of Euler equations is written the identification result of this paper relies on parametric methods of identification of Euler equations.

[13] Note that $\bar{Y}_t$ is not exogenous but it depends on all exogenous variables in each period. It can therefore be varied by changing one of the exogenous variables at $t' \neq t$. In the remainder of the section, it will therefore be treated as an exogenous variable.

the public consumption Euler equation and the public consumption efficiency condition. To understand how these functions can be recovered, note that the public consumption Euler equation can be written in the form

$$f_Q^1\left(\bar{w}_{1t},h_t^1,Q_t\right)+\frac{1-\mu}{\mu}f_Q^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)=$$
$$\beta E_t\left[\left(f_Q^1\left(\bar{w}_{1t+1},h_{t+1}^1,Q_{t+1}\right)+\frac{1-\mu}{\mu}f_Q^2\left(C_{t+1},\bar{w}_{1t+1},h_{t+1}^1,h_{t+1}^2,Q_{t+1}\right)\right)R_{t+1}\frac{P_t}{P_{t+1}}\right],$$

where μ and β are known from the private consumption efficiency condition and Euler equations. The public consumption Euler equation enables one to recover the following function:

$$G\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)=f_Q^1\left(\bar{w}_{1t},h_t^1,Q_t\right)+\frac{1-\mu}{\mu}f_Q^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right). \tag{2.17}$$

Household decisions must also satisfied the following public consumption efficiency condition:

$$\frac{f_Q^1\left(\bar{w}_{1t},h_t^1,Q_t\right)}{f_c^1\left(\bar{w}_{1t},h_t^1,Q_t\right)}+\frac{f_Q^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)}{f_c^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)}=\frac{P_t}{p_t}. \tag{2.18}$$

Note that for every realization of p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$ the functions f_c^1 and f_c^2, and the decision power parameter μ are known. The reduced-form marginal utilities for public consumption can therefore be identified by solving equations (2.17) and (2.18) for f_Q^1 and f_Q^2 for every realization of p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$.

Under the additional assumption that agent 2 chooses to supply a positive amount of labor, agent 2's reduced-form marginal utility of leisure can also be identified by equating her marginal rate of substitution between private consumption and leisure to the real wage, i.e.,

$$\frac{f_l^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)}{f_c^2\left(C_t,\bar{w}_{1t},h_t^1,h_t^2,Q_t\right)}=\bar{w}_{2t}.$$

The function f_c^2 is known from the Euler equations for every realization of the exogenous variables p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$. Thus, f_l^2 can be identified by setting it equal to $\bar{w}_{2t}f_c^2$.

All the reduced-form marginal utilities are therefore identified. However, the information on individual preferences is contained in the original marginal utilities. The following proposition shows that the original marginal utilities are identified if the reduced-form marginal utilities are known and variation in all the exogenous variables is observed.

Proposition 1. If both agents supply a positive amount of labor and either u^1 or u^2 satisfies the invertibility condition (2.14), the marginal utilities u_c^1, u_c^2, u_l^1, u_l^2, u_Q^1, u_Q^2, the decision power μ, and the consumption function g are identified up to the additive constant of g.

If only agent 1 supplies a positive amount of labor and u^1 satisfies the invertibility condition (2.14), all the marginal utilities are identified except the marginal utility of leisure for the spouse that does not work. Moreover, μ and g are identified up to the additive constant of g.

Proof. In the appendix. □

To provide the intuition underlying proposition 1, note that if the function $g\left(\bar{w}_1, h^1, Q\right)$ is known the original marginal utilities can be easily identified by means of equations (2.15) and (2.16). We will now discuss how $g\left(\bar{w}_1, h^1, Q\right)$ can be recovered using variation in variables that are observed in the data. Since some insight for the case of a household with only one worker was provided in the previous section, here we will consider the case of a household in which both parents work. Equation (2.16) implies that for every realization of the exogenous variables agent 2's reduced-form marginal utilities of private and public consumption must satisfy the following identities:

$$f_c^2\left(C, \bar{w}_1, h^1, h^2, Q\right) = u_c^2\left(C - g\left(\bar{w}_1, h^1, Q\right), 1 - h^2, Q\right). \tag{2.19}$$

and

$$f_Q^2\left(C, \bar{w}_1, h^1, h^2, Q\right) = u_Q^2\left(C - g\left(\bar{w}_1, h^1, Q\right), 1 - h^2, Q\right). \tag{2.20}$$

Consider variations in the exogenous variables that generate a group of households with identical $\bar{w}_1$, h_1, h_2, and Q but different C. This group of households enables one to recover $u_{c,c}^2$, i.e. how agent 2's marginal utility of private consumption varies with agent 2's private consumption holding everything else constant. To see this observe that f_c^2 is known, which implies that it is known how f_c^2 varies with C if $\bar{w}_1$, h_1, h_2, and Q are held constant. Since (2.19) is satisfied for every feasible C, how u_c^2 varies with C

holding $\bar{w}_1$, h_1, h_2, and Q constant must be equivalent to how f_c^2 varies with C if $\bar{w}_1$, h_1, h_2, and Q are held constant. Finally, how u_c^2 varies with C holding $\bar{w}_1$, h_1, h_2, and Q constant corresponds to u_{cc}^2. Consequently, $u_{cc}^2 = f_{cc}^2$. The same argument applied to equation (2.20) implies that $u_{Qc}^2 = f_{Qc}^2$.

Consider now changes in the exogenous variables that generate the group of households for which C, $\bar{w}_1$, h_2, and Q are constant, but h_1 varies. This group of households provides joint information on u_{cc}^2 and g_{h_1}. To explain this note that it is known how f_c^2 varies with h_1 if C, $\bar{w}_1$, h_2, and Q are held constant. By (2.19), how u_c^2 varies with h^1 holding C, $\bar{w}_1$, h_2, and Q constant must be equivalent to how f_c^2 varies with h_1 if C, $\bar{w}_1$, h_2, and Q are held constant. Finally, observe that by varying h^1 on the right hand side of (2.19), one obtains information on $u_{cc}^2 g_{h_1}$. This implies that $u_{cc}^2 g_{h_1} = -f_{ch_1}^2$.

Consider the variation in the exogenous variables that generates the group of households for which C, h_1, h_2, and Q are constant, but $\bar{w}_1$ varies. Using the argument employed for the previous group of households, it can be shown that $u_{cc}^2 g_{\bar{w}_1} = -f_{c\bar{w}_1}^2$. Consider the variation in p_t, P_t, w_{1t}, w_{2t}, and $\bar{Y}_t$ that generates the group of households for which C, h_1, h_2, and $\bar{w}_1$ are constant, but Q varies. The logic employed for the previous two groups of households indicates that $-u_{cc}^2 g_Q + u_{cQ}^2 = f_{c\bar{w}_1}^2$.

All the information required to identify how $g\left(\bar{w}^1, h^1, Q\right)$ varies with $\bar{w}_1$, h_1, and Q is now known. Using the first and second group of households one obtains that $g_{h_1} = -f_{ch_1}^2 / f_{cc}^2$. The first and third group of households imply that $g_{\bar{w}_1} = -f_{c\bar{w}_1}^2 / f_{cc}^2$. Using the first and fourth group of households it can be shown that $g_Q = \left(f_{Qc}^2 - f_{c\bar{w}_1}^2\right) / f_{cc}^2$. Finally, since it is known how $g\left(\bar{w}^1, h^1, Q\right)$ varies with $\bar{w}_1$, h_1, and Q, the function g is known up to an additive constant. It is then straightforward to recover the original marginal utilities using the reduced-form marginal utilities and g.

An implication of Proposition 1 is that the individual preferences over private consumption, public consumption, and leisure can be identified. This leads to the following corollary.

Corollary 1. *If both parents work, the individual preferences over public consumption, private consumption, and leisure are identified up to an additive constant.*

If only one parent works, the individual preferences over public and private con-

sumption for both parents and the preferences over leisure for the working parent are identified up to an additive constant.

This Corollary indicates that the preferences for expenditure on children of the mother and father can be identified even if only one of them supplies a positive amount of labor hours. This result should help researchers in predicting which of different policies designed to improve the cognitive and non-cognitive abilities of children will be the most effective.

The implementation of the identification method proposed in this paper requires a longitudinal dataset that contains information on leisure, public consumption, private consumption, and wages. A dataset with these features is the CEX, which is a longitudinal dataset with information on all the required variables. Individual labor supply and wages are observed. Detailed data on expenditure on different consumption items are collected. Moreover, the expenditure data include information on the main components of children expenditure, namely expenditure on children clothing, children shoes, school books, and other educational expenses. All these variables are observed for four consecutive quarters. A drawback of the CEX is that food consumption is only measured at the household level. It is therefore not possible to determine which fraction is consumed by children, which represents public consumption. As a partial solution, the econometrician can either assume that food consumption is separable from other consumption goods and leisure or she can impute the fraction of food items consumed by children using information on the type of goods purchased by the household and the family structure.

This section shows that individual preferences can be identified without assuming a particular utility function. In the following sections, specific utility functions will be used jointly with the identification result presented in this section to estimate the key parameters of the intertemporal collective model.

2.6 Empirical Implementation

The next two subsections will outline the preference and heterogeneity assumptions used in the estimation of individual preferences and the class of measurement errors that are allowed.

2.6.1 Preference Assumptions

The empirical analysis will focus on the estimation of individual preferences for private consumption and leisure. The implicit assumption is that private consumption and leisure are strongly separable from public consumption. Since this assumption is more realistic for the group of households with no children, the estimation will be performed using this restricted sample.

It is assumed that agent i's preferences can be represented using the following utility function:

$$u^i\left(c^i, T-h^i, Q\right) = \frac{\left[\left(c^i\right)^{\sigma_i}\left(T-h^i\right)^{\theta_i}\left(Q\right)^{1-\sigma_i-\theta_i}\right]^{1-\rho_i}}{1-\rho_i},$$

with $\rho_i > 0$, $0 < \sigma_i < 1$, and $0 < \theta_i < 1$. This utility function has been used extensively in the past for its simplicity in research projects that attempt to model the relationship between consumption and leisure. Three notable examples are Kydland and Prescott (1982), Prescott (1986), and Browning, Hansen, and Heckman (1999). The parameter ρ_i captures the intertemporal aspects of individual preferences. In particular, $-1/\rho_i$ is agent i's intertemporal elasticity of substitution, which measures the willingness to substitute the composite good $\left(c^i\right)^{\sigma_i}\left(T-h^i\right)^{\theta_i}\left(Q\right)^{1-\sigma_i-\theta_i}$ between different dates. The parameter σ_i captures the intraperiod features of individual preferences and it measures in each period the fraction of expenditure assigned to agent i which is allocated to private consumption.

The consumption function $g\left(\bar{w}^1, h^1\right)$ corresponding to these preferences can be written in the following form:

$$g\left(\bar{w}^{1}, h^{1}\right)=\frac{\sigma_{1}}{\theta_{1}} \bar{w}_{t}^{1}\left(T-h_{t}^{1}\right).$$

2.6.2 Household Heterogeneity and Measurement Errors

So far the only source of household heterogeneity is the realization of the state of nature. In the estimation of individual preferences, We will allow for two additional sources of heterogeneity. A subscript h will be used to denote an observation for household h.

The main idea underlying the identification of individual preferences is that individual consumption can be written as a function of individual labor supply and own wage. In particular, given the functional form assumed for the utility functions, individual consumption and the individual value of leisure, $\bar{w}_{t}^{1}\left(T-h_{t}^{1}\right)$, should be linearly related and therefore perfectly correlated. This implication of the model can be tested using the sample of singles, since their individual consumption is observed. In the CEX, the correlation is 0.30 for single females and 0.26 for single males. This indicates that there is a positive relationship between individual consumption and value of leisure as predicted by the model. But it also suggests that there is additional heterogeneity characterizing the function g. A potential interpretation of this finding is that, for a given $T-h^{i}$, the perceived value of leisure varies with age, education and seasonal dummies, because the available alternatives vary with these variables. This source of heterogeneity will be captured by assuming that agent i's utility function depends on effective leisure, $\hat{l}^{i}$, where effective leisure is defined as

$$\hat{l}_{t,h}^{i}=\left(T-h_{t,h}^{1}\right) \exp \left(\alpha_{i}^{\prime} z_{t,h}\right),$$

and $z_{t,h}$ is a vector containing the wife's and husband's age, an education dummy for the wife and for the husband, and a seasonal dummy. This implies that[14]

$$c_{t,h}^{1}=g\left(\bar{w}_{t,h}^{1}, h_{t,h}^{1}, z_{t,h}\right)=\frac{\sigma_{1}}{\theta_{1}} \bar{w}_{t,h}^{1}\left(T-h_{t,h}^{1}\right) \exp \left(\alpha_{i}^{\prime} z_{t,h}\right).$$

[14]An alternative interpretation of the low correlation between consumption and value of leisure is that the assumption on preferences is restrictive.

As an additional source of heterogeneity, it will be assumed that the logarithm of the ratio of the Pareto weights varies across households according to an unknown distribution with mean $\overline{l\mu}$. This implies that $\log(\mu_h)$ can be written in the form

$$\log(\mu_h) = \overline{l\mu} + \eta_h,$$

where η is a mean-zero random variable. It is important to remark that under ex-ante efficiency the household ratio of Pareto weights cannot change over time.

Finally, to determine which class of measurement errors can be allowed in the model, we will add measurement errors in private household consumption and individual wages. It will be assumed that the measurement errors satisfy the following three conditions. First, they are additive in the logarithm of private household consumption and individual wages. Second, let $C^*_{t,h}$ and $w^{i*}_{t,h}$ be true private household consumption and wages for household h in period t and denote with $C_{t,h}$ and $w^i_{t,h}$ the observed variables. It is assumed that the true variables can be written in the following form:

$$\log C^*_{t,h} = \log C_{t,h} + \delta_C + \epsilon_{t,h}, \quad \log w^{i*}_{ht} = \log w^i_{t,h} + \delta_{w^i} + \epsilon_{t,h},$$

where δ_C and δ^i_w are two constants and $\epsilon_{t,h}$ is a mean-zero random variable which is common to private consumption and wages.[15] Third, in each period t, the common component $\epsilon_{t,h}$ are independent of the information known to the household.[16]

Let $\gamma_i = \sigma_i(1-\rho_i)$, $\lambda_i = \theta_i(1-\rho_i)$, and $\xi_i = (1-\sigma_i-\theta_i)(1-\rho_i)$. The assumptions on preferences and household heterogeneity imply that the transformed marginal utilities for consumption have the following form:

$$f^1_c = \sigma_1\left(\frac{\sigma_1}{\theta_1}\right)^{\gamma_1-1}\left(\bar{w}^1_{t,h}\right)^{\gamma_1-1}\left(T-h^1_{t,h}\right)^{\gamma_1+\lambda_1-1}(Q_{t,h})^{\xi_1}e^{\left((\gamma_1+\lambda_1-1)\alpha_1 z_{t,h}\right)}e^{(\gamma_1-1)\left(\delta_{w^1}+\epsilon_{t,h}\right)}$$

---

[15]Under the standard assumption that measurement errors have zero mean, the constants δ_C and δ_{w^i} must be equal to zero and the consumption and wage measurement errors must be identical.

[16]Preferences will also be estimated for single agents. In married households two respondents provide information on consumption and wages, whereas in single households only one respondent is present at the interview. To take this into account, in the estimation of preferences for singles we will also allow for measurement errors $\varepsilon^s_{C,t,h}$ and $\varepsilon^s_{w,t,h}$ that are specific to singles.

$$f_c^2 = \sigma_2 \left(C_{t,h} e^{\left(\delta_C + \epsilon_{t,h}\right)} - \frac{\sigma_1}{\theta_1} \bar{w}^1_{t,h} e^{\left(\delta_{w^1} + \epsilon_{t,h}\right)} \left(T - h^1_{t,h}\right) e^{\alpha_1 z_{t,h}} \right)^{\gamma_2 - 1} \left(T - h^2_{t,h}\right)^{\lambda_2} \left(Q_{t,h}\right)^{\xi_2} e^{\alpha_2 \lambda_2 z_{t,h}}.$$

Hence, if one defines $\hat{\phi} = \dfrac{\sigma_1 \exp\left(\delta_{w^1}\right)}{\theta_1 \exp\left(\delta_C\right)}$, the individual Euler equations (2.5) can be written in the following form:

$$\beta_1 E_t \left[\left(\frac{\bar{w}^1_{t+1,h}}{\bar{w}^1_{t,h}} \right)^{\gamma_1 - 1} \left(\frac{T - h^1_{t+1,h}}{T - h^1_{t,h}} \right)^{\gamma_1 + \lambda_1 - 1} \left(\frac{Q_{t+1,h}}{Q_{t,h}} \right)^{\xi_1} e^{(\gamma_1 + \lambda_1 - 1)\alpha_1 \left(z_{t+1,h} - z_{t,h}\right)} R_{t+1,h} \frac{p_t}{p_{t+1}} \right] = \frac{e^{(\gamma_1 - 1)\left(\delta_{w^1} + \epsilon_{t,h}\right)}}{E_t \left[e^{(\gamma_1 - 1)\left(\delta_{w^1} + \epsilon_{t+1,h}\right)} \right]}.$$

The individual Euler equations (2.5) can be written in the following form:

$$\beta_2 E_t \left[\left(\frac{C_{t+1,h} - \hat{\phi} \bar{w}^1_{t+1,h} \left(T - h^1_{t+1,h}\right) e^{\alpha_1 z_{t+1,h}}}{C_{t,h} - \hat{\phi} \bar{w}^1_{t,h} \left(T - h^1_{t,h}\right) e^{\alpha_1 z_{t,h}}} \right)^{\gamma_2 - 1} \left(\frac{T - h^2_{t+1,h}}{T - h^2_{t,h}} \right)^{\lambda_2} \left(\frac{Q_{t+1,h}}{Q_{t,h}} \right)^{\xi_2} \times e^{\alpha_2 \lambda_2 \left(z_{t+1,h} - z_{t,h}\right)} R_{t+1,h} \frac{p_t}{p_{t+1}} \right] = \frac{e^{(\gamma_2 - 1)\left(\delta_C + \epsilon_{t,h}\right)}}{E_t \left[e^{(\gamma_2 - 1)\left(\delta_C + \epsilon_{t+1,h}\right)} \right]}.$$

Finally, if one takes the logarithm of the efficiency equation (2.5), this condition can be written as follows:

$$\begin{gathered}
(\gamma_1 - 1) \log \bar{w}^1_{t,h} + (\gamma_1 + \lambda_1 - 1) \log \left(T - h^1_{t,h}\right) + (\xi_1 - \xi_2) Q_{t,h} - \\
(\gamma_2 - 1) \log \left(C_{t,h} - \hat{\phi} \bar{w}^1_{t,h} \left(T - h^1_{t,h}\right) e^{\alpha_1 z_{t,h}} \right) - \lambda_2 \log \left(T - h^2_{t,h}\right) - \\
\left(\alpha_1 (\gamma_1 + \lambda_1 - 1) + \alpha_2 \lambda_2\right) z_{t,h} = \overline{l\mu} + \eta_h + (\log \sigma_2 - \log \sigma_1) - (\gamma_1 - 1) \log \hat{\phi} + (\gamma_2 - \gamma_1)(\delta_C + \epsilon_{t,h}).
\end{gathered}$$

Taking the unconditional expectation of both sides, agent 1's Euler equations become

$$\beta_1 E \left[\left(\frac{\bar{w}^1_{t+1,h}}{\bar{w}^1_{t,h}} \right)^{\gamma_1 - 1} \left(\frac{T - h^1_{t+1,h}}{T - h^1_{t,h}} \right)^{\gamma_1 + \lambda_1 - 1} \left(\frac{Q_{t+1,h}}{Q_{t,h}} \right)^{\xi_1} e^{(\gamma_1 + \lambda_1 - 1)\alpha_1 \left(z_{t+1,h} - z_{t,h}\right)} R_{t+1,h} \frac{p_t}{p_{t+1}} \right] = 1, \tag{2.21}$$

agent 2's Euler equations can be written as

$$\beta_2 E\left[\left(\frac{C_{t+1,h}-\hat{\phi}\bar{w}^1_{t+1,h}\left(T-h^1_{t+1,h}\right)e^{\alpha_1 z_{t+1,h}}}{C_{t,h}-\hat{\phi}\bar{w}^1_{t,h}\left(T-h^1_{t,h}\right)e^{\alpha_1 z_{t,h}}}\right)^{\gamma_2-1}\left(\frac{T-h^2_{t+1,h}}{T-h^2_{t,h}}\right)^{\lambda_2}\left(\frac{Q_{t+1,h}}{Q_{t,h}}\right)^{\xi_2}\times\right.$$
$$\left. e^{\alpha_2\lambda_2\left(z_{t+1,h}-z_{t,h}\right)}R_{t+1,h}\frac{p_t}{p_{t+1}}\right]=1, \tag{2.22}$$

and the efficiency condition becomes

$$E\left[(\gamma_1-1)\log\bar{w}^1_{t,h}+(\gamma_1+\lambda_1-1)\log\left(T-h^1_{t,h}\right)+(\xi_1-\xi_2)\,Q_{t,h}-\right. \tag{2.23}$$
$$(\gamma_2-1)\log\left(C_{t,h}-\hat{\phi}\bar{w}^1_{t,h}\left(T-h^1_{t,h}\right)e^{\alpha_1 z_{t,h}}\right)-\lambda_2\log\left(T-h^2_{t,h}\right) \tag{2.24}$$
$$\left.-\left(\alpha_1\left(\gamma_1+\lambda_1-1\right)+\alpha_2\lambda_2\right)z_{t,h}\right]=\overline{l\mu}+\left(\log\sigma_2-\log\sigma_1\right)-(\gamma_1-1)\log\hat{\phi}+(\gamma_2-\gamma_1)\,\delta_C. \tag{2.25}$$

Using the CEX data, the coefficients γ_i, λ_i, $\hat{\phi}$, α_i, β_i, and $\overline{l\mu}+(\log\sigma_2-\log\sigma_1)-(\gamma_1-1)\log\hat{\phi}+(\gamma_2-\gamma_1)\,\delta_C$ will be estimated by applying the Generalized Method of Moments (GMM) to these three equations. The remaining parameters of the individual utility functions can then be recovered using the following equations:

$$\gamma_i=\sigma_i\left(1-\rho_i\right),\qquad \lambda_i=\theta_i\left(1-\rho_i\right),\qquad \xi_i=\left(1-\sigma_i-\theta_i\right)\left(1-\rho_i\right).$$

Finally, it is important to remark that the mean of the logarithm of the ratio of the Pareto weights $\overline{l\mu}$ can be identified only if it is assumed that the constant δ_C is equal to zero.

2.7 Econometric Issues

The transformed Euler equations of agent 2 cannot be log-linearized. Consequently, individual preferences will be estimated using the non-linear transformed Euler equations jointly with the efficiency condition. The non-linearities in the Euler equations imply that we can only allow for the class of measurement errors introduced in the previous section, which is a special case of the measurement errors that can be allowed in linear models. To evaluate the effect of the non-linearities on the coefficient estimates, the sample of single households will be used. In particular, the estimation of individual preferences for single households requires only the transformed Euler equation of agent 1, which can be log-linearized. The individual preferences can therefore be estimated using a linear and a non-linear version of the model. The estimates can then be compared to examine the effect of a larger class of measurement errors.

The identification result of Proposition 1 holds only if the consumption function g is well defined. This requires that at least one household member decides to supply a positive amount of labor in two consecutive periods. In the sample of couples used in the estimation the fraction of households in which the husband supplies a positive amount of labor for the entire survey period is around 80%, whereas the fraction in which the wife works during the survey is around 70%. In spite of this, in the estimation the wive's labor supply and wage will be used to derive the consumption function g for the following two reasons. First, there is more variation in the labor supply of married women relative to married men. Second, the correlation between individual consumption and value of leisure is higher for single females relative to single males, which suggests that the correlation should be higher for married women relative to married men. All this implies that the sample used in the estimation can be composed only of households in which the wife works during the survey period. As a consequence, if the residuals of the transformed Euler equations are correlated with the labor force participation decisions of the wife, the estimation results will be affected by a selection bias.

To quantify the selection bias, we will use the sample of single households. Denote with D_t^1 a dummy equal to 1 if agent 1 works in period t and let ζ_{t+1}^i be the error term corresponding to the transformed Euler equations of agent i. Since individual preferences are estimated using the sample of households in which agent 1 works at t and $t+1$, the parameter estimates of both couples and singles are unbiased only if

$$E\left[\zeta_{t+1}^i \,\middle|\, D_t^1 = 1, D_{t+1}^1 = 1\right] = 0,$$

i.e. only if ζ_{t+1}^i is independent of the participation decisions in period t and $t+1$. Suppose this independence assumption is not satisfied. For both couples and singles, the labor force participation decision of agent 1 can be formulated using the model proposed in this paper. In particular, in each period t agent 1's marginal rate of substitution between leisure and consumption is equal to the real wage if $D_t^1 = 1$, but it is greater than the real wage if $D_t^1 = 0$. Under the functional form assumptions for the individual utilities, this implies that

$$\begin{aligned} c_t^1 &\leq \phi \bar{w}_t^1 T \quad \text{if } D_t^1 = 1, \\ c_t^1 &> \phi \bar{w}_t^1 T \quad \text{if } D_t^1 = 0. \end{aligned}$$

It is assumed that the wage equation is determined outside the model and that it can be written as

$$\log w_t^1 = X_t \beta + e_t,$$

where X_t includes labor market experience, its square, and a price index that is household and region specific. The selection equations in period t can then be written in the form

$$\begin{aligned} \log c_t^1 - \log \phi - \log T - X_t \beta &\leq e_t \quad \text{if } D_t^1 = 1, \\ \log c_t^1 - \log \phi - \log T - X_t \beta &> e_t \quad \text{if } D_t^1 = 0. \end{aligned}$$

Suppose that ζ_{t+1}^i, e_t, and e_{t+1} are normally distributed with mean vector 0 and covariance matrix

$$\begin{bmatrix} \sigma_{\zeta^i}^2 & \rho_{\zeta^i,t} & \rho_{\zeta^i,t+1} \\ & 1 & \rho \\ & & 1 \end{bmatrix}.$$

Then, by Tunali (1986),

$$E\left[\zeta_{t+1}^i \left| D_t^1 = 1, D_{t+1}^1 = 1\right.\right] = \sigma_{\zeta^i} \rho_{\zeta^i,t} \xi_t^i + \sigma_{\zeta^i} \rho_{\zeta^i,t+1} \xi_{t+1}^i,$$

where

$$\xi_t = \frac{\phi\left(X_t\beta\right) \Phi\left(\dfrac{X_{t+1}\beta - \rho X_t \beta}{\left(1-\rho^2\right)^{1/2}}\right)}{G\left(X_t\beta, X_{t+1}\beta, \rho\right)}, \quad \xi_{t+1} = \frac{\phi\left(X_{t+1}\beta\right) \Phi\left(\dfrac{X_t\beta - \rho X_{t+1} \beta}{\left(1-\rho^2\right)^{1/2}}\right)}{G\left(X_t\beta, X_{t+1}\beta, \rho\right)},$$

and where ϕ, Φ and G are, respectively, the standard univariate normal density function, the standard univariate normal distribution function, and the standard bivariate normal distribution function. Individual consumption is observed for singles without children. Consequently, for this group of households the individual Euler equations can be estimated jointly with the labor force participation equations to quantify the selection bias. Following Newey and McFadden (1994), the Euler equations adjusted for selection are estimated using GMM in one step by adding as moment conditions the first order conditions of the bivariate probit, which determines the probability of being in one of the four possible labor supply states defined by D_t^1 and D_{t+1}^1. A similar approach could be used for couples, but stronger assumptions are required.[17]

The residuals of the individual Euler equations contain the expectation error implicit in these intertemporal optimality conditions. Since part of the expectation error is generated by aggregate shocks, it could be correlated across households. As suggested by Chamberlain (1984), this implies that the Euler equations can be consistently estimated only if the sample period covered by the data is long enough to contain all the stages of the business cycle. For this reason, data from 1982 to 1998 are used in the estimation.

The Euler equations and the efficiency conditions will be estimated using the continuous updating GMM. The choice of this GMM estimator is based on work by Hansen, Heaton, and Yaron (1996) and Donald and Newey (2000) indicating that the continuous updating GMM estimator has smaller bias than the more common two-step efficient GMM estimator, with and without autocorrelation. Under the assumption of rational expectations, any variable known at time t should be a valid instrument for GMM. The existence of measurement errors, however, may introduce dependence between variables known at time t and concurrent and future variables, even under rational expectations.

[17]To estimate agent 1's participation equations for couples, individual consumption must be substituted out using the first order conditions for consumption, which depend on the budget constraint multiplier in the corresponding period. Using the Euler equations and an approach similar to Heckman and MaCurdy (1980) and Browning et al. (1986), the multiplier in each period can be written as a function of the multiplier at 0 and the sequence of interest rates. If a long panel is available, the participation equations can therefore be estimated jointly with the individual Euler equations and efficiency conditions by using a fixed effect estimator. Unfortunately, the panel used in this paper covers only two consecutive period, which implies that the participation equations can be estimated only if it is assumed that the initial multiplier is constant across households or by using as a proxy for the multiplier initial wealth.

To address this problem, only variables known at $t-1$ are used. This requires three consecutive observations for the same household: two to compute the growth rate for consumption, leisure, and wages, and at least one additional observation to construct the instrument set. In the CEX, labor supply and labor income data are only measured in the first and last interview, which implies that only two consecutive observations are available for each household. To address this problem, the set of instruments is constructed employing lagged cohort variables, where the cohort variables are computed using 7-years intervals for the head's year of birth.

2.8 CEX Data

The CEX survey is a rotating panel organized by the Bureau of Labor Statistics (BLS). Each quarter about 4,500 households, representative of the U.S. population, are interviewed: 80 percent are reinterviewed the following quarter, while the remaining 20 percent are replaced by a new randomly selected group. Each household is interviewed at most for four quarters and detailed information is collected on expenditures, income, and demographics. Following Meghir and Weber (1996) household level data for the available quarters are used in the estimation. The sample employed in this paper covers the period 1982-1998. The first two years are excluded because the data were collected with a slightly different methodology.

The CEX collects consumption data in each quarter of the survey. Labor supply and labor income data, however, are gathered only during the first and last interviews unless a member of the household reports changing his or her employment. In the second and third interviews the labor variables are set equal to the data reported in the first interview. Consequently, in the estimation I use quarterly variables computed using the first and last interviews.

Quarterly household consumption of singles is computed as the sum of food at home, food away from home, tobacco, alcohol, public and private transportation, personal care, clothing, house maintenance, heating fuel, utilities, housekeeping services, and transportation repairs, which is the definition used in Attanasio and Weber (1995). Household consumption of couples is obtained by subtracting the expenditure on goods that are clearly public consumption from the definition used for singles, namely house maintenance, heating fuel, and housekeeping services. Quarterly individual labor supply is calculated as the number of hours usually worked per week multiplied by 13 weeks. The total amount of time that an agent can divide between labor supply and leisure, T, is set equal to 1183, which is equal to 13 hours per day times 7 days a week times 13 weeks a quarter.[18] Quarterly leisure can then be computed as T minus quarterly labor supply. The individual hourly wage rate is determined using three variables: the amount of the last gross pay, the time period the last gross pay covered, and the number of hours usually worked per week in the corresponding period. The after-tax wage rate is computed using federal effective tax rates generated by the NBER's TAXSIM model. The gross interest rate is obtained compounding the 20-year municipal bond rate for the three quarters that separate the first interview from the last. Household consumption, individual after-tax wages, and the gross interest rate are deflated using a household specific price index. The index is calculated as a weighted average of the consumer price indices published by the Bureau of Labor Statistics, with weights equal to the expenditure share for the particular consumption good.

The identification result requires that at least one household member supplies a positive amount of labor in two consecutive periods. Consequently, we drop from the sample couples in which the wife does not work during at least one of the two quarters used in the estimation. For singles, we drop a household if the head does not work in one of the two quarters. Households with children and households in which the head is older than 65 and younger than 22 are also excluded. Households with missing values in one of the variables defining the individual Euler equations and efficiency condition are dropped. For couples, a household is not used in the estimation if the husband's or the wife's labor supply is lower than 20 hours, or the wife's real after-tax hourly wage is larger than 50 dollars. For singles, we drop a household if the head's labor supply is less than 20 hours or the real after-tax hourly wage is larger than 50 dollars.[19] Summary statistics in 1984 dollars for the main variables are reported in Table 2.1.

[18]The 13 hours per day are computed by allocating 8 hours to sleep, 1 hour to the time required to reach the workplace, and 2 hours to exogenous household production. We also experimented with 12 and 14 hours per day. This change has a small effect on the estimation of σ, which can be explained by noting that, for any level of labor supply, T determines the amount of leisure. However, the main findings of the paper do not change. An alternative approach would be to use a time survey to compute T for married females and males, and for single females and males.

[19]The fraction of couples in which the wife's wage is larger than 50 dollars is around 0.5 percent. The fraction of single males and females is around 0.2 percent. The fraction of couples in which the wife works less than 20 hours is 5 percent of the sample. The fraction of singles in which the head works less than 20 hours is around 1 percent for males and around 2 percent for females.

Table 2.1: Summary Statistics

Independent Variable	Mean for Singles	Mean for Couples
Real Consumption per Quarter	1563.8	2545.1
Head's Labor Supply per Week	42.7	44.4
Spouse's Labor Supply per Week	-	32.1
Conditional Spouse's Labor Supply per Week	-	38.8
Head's After Tax Wage per Hour	7.7	9.2
Wife's Before Tax Wage per Hour	-	6.6
Number of Observations	9464	5064
Number of Families	2366	1266

2.9 Results

To evaluate the performance of the identification result, individual preferences are initially estimated for single agents using several specifications. First, preferences are estimated using standard household consumption Euler equations and the intraperiod condition. Under the assumptions on preferences and heterogeneity of section 2.6, the two equations can be written as follows:

$$E\left[\left(\frac{C_{t+1,h}}{C_{t,h}}\right)^{\gamma_1-1}\left(\frac{T-h_{t+1,h}}{T-h_{t,h}}\right)^{\lambda_1}\left(\frac{Q_{t+1,h}}{Q_{t,h}}\right)^{\xi_1} e^{\alpha\theta\left(z_{t+1,h}-z_{t,h}\right)}\beta R_{t+1,h}\frac{p_{t,h}}{p_{t+1,h}}\right]=1, \tag{2.26}$$

$$E\left[\log C_{t,h}-\log\hat{\phi}-\log\left(\bar{w}^1_{t,h}\left(T-h^1_{t,h}\right)\right)-\alpha' z_{t,h}\right]=0. \tag{2.27}$$

This specification corresponds to the approach traditionally used by the intertemporal literature, except that the intraperiod condition is included in the estimation to pin down the intraperiod parameter σ.[20] Second, preferences of singles are estimated using agent 1's transformed consumption Euler equations (2.21) and the identification result. Note that the transformed Euler equations (2.21) contain the same information as equations (2.26) and (2.27), since they are obtained by substituting the intraperiod condition in the standard Euler equations.

The standard estimation and the estimation based on the identification result will be implemented using log-linearized as well as non-linear Euler equations. All specifications are estimated with and without selection correction terms.

The results for the log-linearized version of the model are reported in Table 2.4 for

[20] Theoretically, σ can be estimated using only the standard household consumption Euler equations. But empirically σ can be precisely estimated only if the intraperiod condition is added to the estimation.

females and Table 2.5 for males. The estimates obtained using the identification result are similar to the ones obtained using the standard method. The estimates for the coefficient ρ are about 1.7 for single males and 5 for single females. The parameter σ is precisely estimated only if the intraperiod condition is added to the estimation as an additional moment condition. In this case the wife's σ is around 0.15, whereas the husband's is around 0.50. Note that in the estimation of couples' preferences an intraperiod condition will be used in the form of the efficiency equation. This will enable me to precisely estimate σ.

The results obtained using the non-linear version of the model are reported in Table 2.6 for single females and Table 2.7 for single males and are similar to the estimates obtained using the log-linearized Euler equations. This suggests that the estimation results do not vary if measurement errors and unobserved heterogeneity are generalized to the class that can be allowed in linear models. The addition of the selection terms to the model does not produce significant differences in the results, which are reported in Table 2.8, Table 2.9, Table 2.10, Table 2.11. In all specifications the selection terms are never statistically significant. This finding can be interpreted in two different ways. Either we are not able to precisely estimate the labor force participation decision, or the unobservable heterogeneity in the participation decision is independent of the Euler equation error term. In most specifications both experience and its square have a statistically significant effect on the participation decision. This suggests that the second interpretation is plausible and that selection biases should not have significant effects on the estimation of individual preferences for couples.

The main empirical results are the estimates for couples, which are obtained using the transformed Euler equation (2.21) for the wife, the transformed Euler equation (2.22) for the husband, and the efficiency condition (2.25). The results are reported in Table 2.12. The wife's ρ is estimated to be around 4.4, whereas the husband's ρ is estimated to be around 2.5. Moreover, the difference between the wife's estimated ρ and the husband's is statistically significant. The wife's and husband's σ are estimated to be around 0.28. Table 2.12also reports the estimate of the mean relative decision power under the assumption that the constant δ_C in the measurement errors is equal to zero. Note that the estimation of the model produces an estimate of the logarithm of the expected value of μ. The reported estimate is obtained by taking a first order Taylor expansion of it. The mean relative decision power is measured to be around 0.82, but the standard errors are five times as large. The last column of Table 2.12 reports the results of the estimation of the standard unitary model with separability between consumption and leisure. The estimate of ρ is 3.7, which is between the estimated ρ for married females and married males and within the range of estimates obtained in the past.

Finally, to test the assumption of ex-ante efficiency, the individual Euler equations (2.21) and (2.22) are also estimated without including the efficiency condition (2.25). Using a distance statistic test with one degree of freedom, ex-ante efficiency cannot be rejected at any standard significance level.

To understand which features of the data generate the differences in intertemporal elasticities of substitution between females and males, consider a single agent. The parameter σ measures the consumption budget share of this individual. In the CEX, the average consumption budget share is around 0.25 for both single females and males. These numbers are consistent with the estimates obtained in this paper which are between 0.12 and 0.55.

To determine how ρ is identified, for a given σ define the composite good $\bar{C} = c^{\sigma} l^{1-\sigma}$. Note that if a single agent decides to save one unit of $\bar{C}$ in period t, she will be able to increase consumption at $t+1$ by $R_{t+1} \left(p_t / p_{t+1}\right)^{\sigma} \left(w_t / w_{t+1}\right)^{1-\sigma}$. We can therefore interpret $R_{t+1} \left(p_t / p_{t+1}\right)^{\sigma} \left(w_t / w_{t+1}\right)^{1-\sigma}$ as the gross return on $\bar{C}$, where R_{t+1} captures the return for investing one unit of $\bar{C}$ in the risk-free asset, and $\left(p_t / p_{t+1}\right)^{\sigma}$ and $\left(w_t / w_{t+1}\right)^{1-\sigma}$ measure the change in prices between t and $t+1$ of the two goods that form $\bar{C}$, weighted using the corresponding budget shares. Using the log-linearized model, it is straightforward to show that $1/\rho$ measures the percentage change in $\bar{C}_{t+1} / \bar{C}_t$ generated by a one percent increase in $R_{t+1} \left(p_t / p_{t+1}\right)^{\sigma} \left(w_t / w_{t+1}\right)^{1-\sigma}$. This elasticity can be determined in the CEX by implementing an IV regression of the logarithm of $\bar{C}_{t+1} / \bar{C}_t$ on the logarithm of $R_{t+1} \left(p_t / p_{t+1}\right)^{\sigma} \left(w_t / w_{t+1}\right)^{1-\sigma}$. If σ is set equal to the average consumption budget share, the estimated coefficient is 0.16 for single females and 0.56 for single males, which explains the estimated heterogeneity in intertemporal preferences.[21]

It is now straightforward to understand how the data generates a different ρ for males and females. Consider two identical single households except that the first one has a female head whereas the second one has a male head. Since males and females have identical σ, these two households face the same return on $\bar{C}$ and given this return they choose $\bar{C}_{t+1} / \bar{C}_t$ optimally. Consider an increase in the rate of return. In the data both single females and males increase the ratio $\bar{C}_{t+1} / \bar{C}_t$. However, the increase in period $t+1$ consumption relative to period t is larger for males, which suggests that females have a higher willingness to pay for a smooth consumption path.

Two remarks are in order. First, there is weak evidence of selection in the marriage market based on individual preferences. In particular, agents that are at the extremes

[21] we use an IV regression instead of an OLS regression to replicate the GMM estimation and to take into consideration that labor supply is used to construct the dependent variable $\bar{C}$ as well as the regressor $R_{t+1} \left(p_{t+1} / p_t\right)^{\sigma} \left(w_{t+1} / w_t\right)^{1-\sigma}$.

of the risk aversion distribution are less likely to be married. Second, according to the results, the elasticity of intertemporal substitution for males is around twice the elasticity for females. Since in the proposed model the parameter ρ is the coefficient of relative risk aversion for the composite good $\left(c^i\right)^{\sigma_i}\left(T-h^i\right)^{1-\sigma_i}$, the results also imply that females are more risk averse than males. In Mazzocco (2007) it is shown that the standard unitary model represents a good approximation of household intertemporal behavior if and only if the individual preferences satisfy a generalization of Gorman aggregation to an intertemporal framework. The heterogeneity in the estimated ρ implies that Gorman aggregation is not satisfied. Consequently, simulations of competing policies based on the standard unitary model are generally misleading, because they do not consider the full extent of intrahousehold risk sharing and specialization that can be obtained if individual preferences are heterogeneous.

One example is the evaluation of the adequacy of household saving at the time of retirement. As shown in Mazzocco (2004), the effect of risk sharing on household saving can be divided into two parts. First, individual members pool their earnings and consequently eliminate part of the uncertainty faced by the household. Under convex marginal utilities, income pooling always has the intuitive effect of reducing saving. Second, household members insure each other by allocating pooled income according to individual risk preferences and decision power. This insurance component of risk sharing can have the counterintuitive effect of raising household saving. The heterogeneity in risk aversion reported in this paper indicates that the insurance component explains a significant fraction of the accumulation and reduction of household wealth. However, as shown in Mazzocco (2004), the unitary model, and therefore any simulation based on it, completely ignores this component of risk sharing. The traditional justification for using the unitary model in simulations in spite of this drawback is that there are no estimates that can be used to fix the parameters that characterize the individual intertemporal preferences. The estimates provided in this paper fill this void.

2.10 Conclusions

In this paper it is shown that the preferences of each decision maker in the household can be identified and estimated even if individual consumption is not observed. The main finding is that there is a significant difference in individual preferences, with the wife exhibiting a greater desire for smooth consumption.

The main implication of this result is that intertemporal decisions cannot be analyzed using a unique utility function for the entire household, because this approach ignores important aspects of intra-household risk sharing and specialization. This implies that any policy analysis related to household intertemporal decisions should be implemented by characterizing each household member by means of individual preferences.

The analysis can be extended in at least one directions. In this paper it is assumed that the time devoted to household production is exogenously given. Under this assumption, it can be incorporated in the available time T. An important project which is left for future research is to generalize the identification result to an environment that allows for endogenous choices of domestic labor. In the meanwhile, empirical works should model T as a function of exogenous variables that determine domestic labor. In this way, differences across households in domestic labor are captured by the heterogeneity in T.

Appendix A: Proofs

Lemma 1

The following Lemma determines the condition under which the marginal rate of substitution function q can be inverted and therefore the consumption function g is well-defined.

Lemma 1. *The function $g\left(\bar{w}_{1t}, h_t^1, Q_t\right)$ is well-defined if*

$$u_{lc}^1\left(c_t^1, 1-h_t^1, Q_t\right) u_c^1\left(c_t^1, 1-h_t^1, Q_t\right) - u_{cc}^1\left(c_t^1, 1-h_t^1, Q_t\right) u_l^1\left(c_t^1, 1-h_t^1, Q_t\right) \neq 0,$$

for any realization of the exogenous variables.

Proof. For any realization of the exogenous variables define

$$d^1\left(c^1, h^1, Q, \bar{w}_1\right) = q^1\left(c_t^1, h_t^1, Q\right) - \bar{w}_{1t} = 0.$$

By the implicit function theorem, $g^1\left(\bar{w}_1, h^1, Q\right)$ is well-defined if $\dfrac{\partial d^1}{\partial c^1} \neq 0$. Which implies the result. □

Proof of Proposition 1

In the second stage of the household problem, the household chooses optimal consumption and leisure in each period and state of nature given $w_{1,t,\omega}$, $w_{2,t,\omega}$, $p_{t,\omega}$, $P_{t,\omega}$, and $\bar{Y}_{t,\omega}$ according to the following problem:

$$\max_{c_{t,\omega}^1, c_{t,\omega}^2, l_{t,\omega}^1, l_{t,\omega}^2, Q_{t,\omega}} \mu u^1(c_{t,\omega}^1, l_{t,\omega}^1, Q_{t,\omega}) + (1-\mu)\, u^2(c_{t,\omega}^2, l_{t,\omega}^2, Q_{t,\omega})$$

$$s.t. \sum_{i=1}^{2}\left(p_{t,\omega} c_{t,\omega}^i + w_{i,t,\omega} l_{t,\omega}^i\right) + P_{t,\omega} Q_{t,\omega} \leq \bar{Y}_{t,\omega}$$

The price of the private good, $p_{t,\omega}$, the price of the public good, $P_{t,\omega}$, agent 1's wage, $w_{1,t,\omega}$, and agent 2's wage, $w_{2,t,\omega}$, represent four independent sources of exogenous variation. The fifth source of variation is $\bar{Y}_{t,\omega}$. It is important to remark that $\bar{Y}_{t,\omega}$ is endogenously determined and it is a function of the exogenous variables in any period and state of nature. This has two implications. First, a change in one of the exogenous variables at $t' \neq t$ and $\omega' \neq \omega$ varies $\bar{Y}_{t,\omega}$. Second, a change in an exogenous variable at $t' \neq t$ and $\omega' \neq \omega$ can vary household decisions in period t and state ω only through $\bar{Y}_{t,\omega}$. In the remainder of the proof a change in $\bar{Y}_{t,\omega}$ should be interpreted as a change in an exogenous variable in period t' and state ω' that varies $\bar{Y}_{t,\omega}$.

Consider first the case in which both agents work. Note that if the function $g\left(\bar{w}_{1t}, h_t^1, Q_t\right)$ can be identified, the original marginal utilities can also be identified by means of the reduced-form marginal utilities which are known. In the remainder of the proof it will be shown that $\frac{\partial g}{\partial \bar{w}_1}$, $\frac{\partial g}{\partial h^1}$, and $\frac{\partial g}{\partial Q}$ can be identified, which implies that $g\left(\bar{w}_1, h^1, Q\right)$ can be identified up to an additive constant.

Consider an arbitrary period t and state ω. Given w_1, w_2, p, P, and $\bar{Y}$, optimal household private consumption, public consumption, agent 1's labor supply, and agent 2's labor supply can be written in the following form:

$$C = C\left(w_1, w_2, p, P, \bar{Y}\right), Q = Q\left(w_1, w_2, p, P, \bar{Y}\right), h^1 = h^1\left(w_1, w_2, p, P, \bar{Y}\right), h^2 = h^2\left(w_1, w_2, p, P, \bar{Y}\right).$$

Agent 2's reduced-form marginal utilities of private consumption and public consumption are defined as follows:

$$f_c^2\left(C, \bar{w}_1, h^1, h^2, Q\right) = u_c^2\left(C - g\left(\bar{w}_1, h^1, Q\right), 1 - h^2, Q\right). \tag{2.28}$$

$$f_Q^2\left(C, \bar{w}_1, h^1, h^2, Q\right) = u_Q^2\left(C - g\left(\bar{w}_1, h^1, Q\right), 1 - h^2, Q\right). \tag{2.29}$$

By construction these equations are satisfied for any combination of w_1, w_2, p, P, and $\bar{Y}$. Consider an arbitrary w_1, w_2, p, P, and $\bar{Y}$. Let dw_1, dw_2, dp, dP, and $d\bar{Y}$ be a small change in the exogenous variables with the following properties: (i) $dw_1 = \frac{w_1}{p}dp$, which implies that $d\bar{w}_1 = 0$; (ii) dw_2, dp, dP, and $d\bar{Y}$ are the solution of the following linear system:

$$\begin{aligned}
\frac{\partial C}{\partial w_2}dw_2 + \left(\frac{\partial C}{\partial w_1}\frac{w_1}{p} + \frac{\partial C}{\partial p}\right)dp + \frac{\partial C}{\partial P}dP + \frac{\partial C}{\partial \bar{Y}}d\bar{Y} &= dC \neq 0,\\
\frac{\partial Q}{\partial w_2}dw_2 + \left(\frac{\partial Q}{\partial w_1}\frac{w_1}{p} + \frac{\partial Q}{\partial p}\right)dp + \frac{\partial Q}{\partial P}dP + \frac{\partial Q}{\partial \bar{Y}}d\bar{Y} &= dQ = 0,\\
\frac{\partial h^1}{\partial w_2}dw_2 + \left(\frac{\partial h^1}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^1}{\partial p}\right)dp + \frac{\partial h^1}{\partial P}dP + \frac{\partial h^1}{\partial \bar{Y}}d\bar{Y} &= dh^1 = 0,\\
\frac{\partial h^2}{\partial w_2}dw_2 + \left(\frac{\partial h^2}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^2}{\partial p}\right)dp + \frac{\partial h^2}{\partial P}dP + \frac{\partial h^2}{\partial \bar{Y}}d\bar{Y} &= dh^2 = 0,
\end{aligned}$$

i.e., the change varies household private consumption, but household public consumption, agent 1's labor supply, and agent 2's labor supply stay constant. The change in f_c^2 implied by dw_1, dw_2, dp, dP, and $d\bar{Y}$ can be computed as follows:[22]

$$df_c^2 = \frac{\partial f_c^2}{\partial C}dC + \frac{\partial f_c^2}{\partial \bar{w}_1}d\bar{w}_1 + \frac{\partial f_c^2}{\partial h^1}dh^1 + \frac{\partial f_c^2}{\partial h^2}dh^2 + \frac{\partial f_c^2}{\partial Q}dQ = \frac{\partial f_c^2}{\partial C}dC.$$

Similarly, the change in u_c^2 implied by dw_1, dw_2, dp, dP, and $d\bar{Y}$ can be written in the following form:

$$du_c^2 = -\frac{\partial u_c^2}{\partial c^2}dC.$$

Since equation (2.28) is satisfied for any w_1, w_2, p, P, and $\bar{Y}$, the change in f_c^2 must equal the change in u_c^2. Consequently,

$$\frac{\partial f_c^2}{\partial C} = -\frac{\partial u_c^2}{\partial c^2}.$$

Since $\dfrac{\partial f_c^2}{\partial C}$ is known, $\dfrac{\partial u_c^2}{\partial c^2}$ is also known.

Consider a change dw_1, dw_2, dp, dP, and $d\bar{Y}$ with the following properties: (i) $dw_1 = \dfrac{w_1}{p}dp$, which implies that $d\bar{w}_1 = 0$; (ii) dw_2, dp, dP, and $d\bar{Y}$ are the solution of

[22] Alternatively, one could totally differentiate f_c^2 with respect to the exogenous variables w_1, w_2, p, P, and $\bar{Y}$ and then impose the constraints implied by the system of linear equations.

the following linear system:

$$\begin{aligned}
\frac{\partial C}{\partial w_2}dw_2 + \left(\frac{\partial C}{\partial w_1}\frac{w_1}{p} + \frac{\partial C}{\partial p}\right)dp + \frac{\partial C}{\partial P}dP + \frac{\partial C}{\partial \bar{Y}}d\bar{Y} = dC = 0,\\
\frac{\partial Q}{\partial w_2}dw_2 + \left(\frac{\partial Q}{\partial w_1}\frac{w_1}{p} + \frac{\partial Q}{\partial p}\right)dp + \frac{\partial Q}{\partial P}dP + \frac{\partial Q}{\partial \bar{Y}}d\bar{Y} = dQ = 0,\\
\frac{\partial h^1}{\partial w_2}dw_2 + \left(\frac{\partial h^1}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^1}{\partial p}\right)dp + \frac{\partial h^1}{\partial P}dP + \frac{\partial h^1}{\partial \bar{Y}}d\bar{Y} = dh^1 \neq 0,\\
\frac{\partial h^2}{\partial w_2}dw_2 + \left(\frac{\partial h^2}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^2}{\partial p}\right)dp + \frac{\partial h^2}{\partial P}dP + \frac{\partial h^2}{\partial \bar{Y}}d\bar{Y} = dh^2 = 0,
\end{aligned}$$

i.e., the change varies agent 1's labor supply, but household private consumption, public consumption, and agent 2's labor supply stay constant. According to equation (2.28), the implied change in f_c^2 must equal the implied change in u_c^2. Consequently, the following equation must be satisfied:[23]

$$\frac{\partial f_c^2}{\partial h^1} = -\frac{\partial u_c^2}{\partial c^2}\frac{\partial g}{\partial h^1}.$$

Since $\dfrac{\partial f_l^2}{\partial h^1}$ and $\dfrac{\partial u_c^2}{\partial c^2}$ are known, $\dfrac{\partial g}{\partial h^1}$ is identified.

Consider a change dw_1, dw_2, dp, dP, and $d\bar{Y}$ with the following properties: (i) $dw_1 \neq \dfrac{w_1}{p}dp$, which implies that $d\bar{w}_1 \neq 0$; (ii) dw_1, dw_2, dp, dP, and $d\bar{Y}$ are the solution of the following linear system:

$$\begin{aligned}
\frac{\partial C}{\partial w_1}dw_1 + \frac{\partial C}{\partial w_2}dw_2 + \frac{\partial C}{\partial p}dp + \frac{\partial C}{\partial P}dP + \frac{\partial C}{\partial \bar{Y}}d\bar{Y} = dC = 0,\\
\frac{\partial Q}{\partial w_1}dw_1 + \frac{\partial Q}{\partial w_2}dw_2 + \frac{\partial Q}{\partial p}dp + \frac{\partial Q}{\partial P}dP + \frac{\partial Q}{\partial \bar{Y}}d\bar{Y} = dQ = 0,\\
\frac{\partial h^1}{\partial w_1}dw_1 + \frac{\partial h^1}{\partial w_2}dw_2 + \frac{\partial h^1}{\partial p}dp + \frac{\partial h^1}{\partial P}dP + \frac{\partial h^1}{\partial \bar{Y}}d\bar{Y} = dh^1 = 0,\\
\frac{\partial h^2}{\partial w_1}dw_1 + \frac{\partial h^2}{\partial w_2}dw_2 + \frac{\partial h^2}{\partial p}dp + \frac{\partial h^2}{\partial P}dP + \frac{\partial h^2}{\partial \bar{Y}}d\bar{Y} = dh^2 = 0,
\end{aligned}$$

i.e., the change does not vary household private consumption, public consumption, agent 1's labor supply, and agent 2's labor supply. By equation (2.28), the implied change in f_c^2 must equal the implied change in u_c^2, which implies that the following

[23]The steps used to derive this equation are equivalent to the steps used to derive (2.10).

equation must be satisfied:

$$\frac{\partial f_c^2}{\partial \bar{w}_1} = -\frac{\partial u_c^2}{\partial c^2}\frac{\partial g}{\partial \bar{w}_1}.$$

Since $\frac{\partial f_c^2}{\partial \bar{w}_1}$ and $\frac{\partial u_c^2}{\partial c^2}$ are known, $\frac{\partial g}{\partial \bar{w}_1}$ is identified.

Similarly since the implied change in f_c^2 and f_Q^2 must equal the implied change in, respectively, u_c^2 and u_Q^2, the following equations must be satisfied:

$$\frac{\partial f_c^2}{\partial \bar{w}_1} = -\frac{\partial u_c^2}{\partial c^2}\frac{\partial g}{\partial \bar{w}_1},$$

$$\frac{\partial f_Q^2}{\partial \bar{w}_1} = -\frac{\partial u_Q^2}{\partial c^2}\frac{\partial g}{\partial \bar{w}_1}.$$

Note that $\frac{\partial f_c^2}{\partial \bar{w}_1}$, $\frac{\partial f_Q^2}{\partial \bar{w}_1}$, and $\frac{\partial g}{\partial \bar{w}_1}$ are known, which implies that $\frac{\partial u_Q^2}{\partial c^2}$ and $\frac{\partial u_c^2}{\partial c^2}$ are identified.

Finally, consider a change dw_1, dw_2, dp, dP, and $d\bar{Y}$ with the following properties: (i) $dw_1 = \frac{w_1}{p}dp$, which implies that $d\bar{w}_1 = 0$; (ii) dw_2, dp, dP, and $d\bar{Y}$ are the solution of the following linear system:

$$\begin{aligned}
\frac{\partial C}{\partial w_2}dw_2 + \left(\frac{\partial C}{\partial w_1}\frac{w_1}{p} + \frac{\partial C}{\partial p}\right)dp + \frac{\partial C}{\partial P}dP + \frac{\partial C}{\partial \bar{Y}}d\bar{Y} &= dC = 0,\\
\frac{\partial Q}{\partial w_2}dw_2 + \left(\frac{\partial Q}{\partial w_1}\frac{w_1}{p} + \frac{\partial Q}{\partial p}\right)dp + \frac{\partial Q}{\partial P}dP + \frac{\partial Q}{\partial \bar{Y}}d\bar{Y} &= dQ \neq 0,\\
\frac{\partial h^1}{\partial w_2}dw_2 + \left(\frac{\partial h^1}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^1}{\partial p}\right)dp + \frac{\partial h^1}{\partial P}dP + \frac{\partial h^1}{\partial \bar{Y}}d\bar{Y} &= dh^1 = 0,\\
\frac{\partial h^2}{\partial w_2}dw_2 + \left(\frac{\partial h^2}{\partial w_1}\frac{w_1}{p} + \frac{\partial h^2}{\partial p}\right)dp + \frac{\partial h^2}{\partial P}dP + \frac{\partial h^2}{\partial \bar{Y}}d\bar{Y} &= dh^2 = 0,
\end{aligned}$$

i.e., the change varies public consumption, but it does not vary household private consumption, agent 1's labor supply, and agent 2's labor supply. According to (2.28), the implied change in f_c^2 must equal the implied change in u_c^2, which implies that

$$\frac{\partial f_c^2}{\partial Q} = -\frac{\partial u_c^2}{\partial c^2}\frac{\partial g}{\partial Q} + \frac{\partial u_c^2}{\partial Q}.$$

Observe that $\frac{\partial f_c^2}{\partial Q}$, $\frac{\partial u_Q^2}{\partial c^2}$, and $\frac{\partial u_c^2}{\partial c^2}$ are known. Consequently, $\frac{\partial g}{\partial Q}$ is identified.

Since $\dfrac{\partial g}{\partial h^1}$, $\dfrac{\partial g}{\partial \bar{w}_1}$, and $\dfrac{\partial g}{\partial Q}$ are known, the function g is identified up to the constant of integration. It is then straightforward to use $g(\bar{w}_1, h^1, Q)$ to recover u_c^i, u_l^i, and u_Q^i from f_c^i, f_l^i, and f_Q^i up to the additive constant of g.

It is important to remark that the proof requires that the following matrix of coefficients of the linear systems is of full rank:

$$\begin{bmatrix} \dfrac{\partial C}{\partial w_2} & \dfrac{\partial C}{\partial p} & \dfrac{\partial C}{\partial P} & \dfrac{\partial C}{\partial \bar{Y}} \\ \dfrac{\partial Q}{\partial w_2} & \dfrac{\partial Q}{\partial p} & \dfrac{\partial Q}{\partial P} & \dfrac{\partial Q}{\partial \bar{Y}} \\ \dfrac{\partial h^1}{\partial w_2} & \dfrac{\partial h^1}{\partial p} & \dfrac{\partial h^1}{\partial P} & \dfrac{\partial h^1}{\partial \bar{Y}} \\ \dfrac{\partial h^2}{\partial w_2} & \dfrac{\partial h^2}{\partial p} & \dfrac{\partial h^2}{\partial P} & \dfrac{\partial h^2}{\partial \bar{Y}} \end{bmatrix}.$$

There are two cases in which this condition is not satisfied: (i) at least one of the demand functions is independent of all the exogenous variables; (ii) the rows or columns are linearly dependent. Since the first case is not realistic, I will only discuss the second one. The rows of the matrix are linearly dependent if the variation in one of the demand functions generated by changes in the exogenous variables provides no additional information conditional on the variation in the other demand functions. The columns are linearly dependent if a change in one of the exogenous variables provide no additional information on how the demand functions C, Q, h^1, and h^2 vary conditional on the variation generated by the other exogenous variables. This emphasizes that the identification of individual preferences requires that independent variations in C, Q, h^1, and h^2 are observed and that the exogenous variables can generate it.

Consider the case in which only agent 1 supplies a positive amount of labor. In this case, h^2 is always equal to zero, no variation in w_2 is observed, and the reduced-form marginal utility f_l^2 is not known. In the first part of the proof, the equation defining f_l^2 and variation in h^2 were never used. Consequently, the previous argument can also be applied to households in which only one agent supplies a positive amount of labor by dropping h^2 from equations (2.28) and (2.29), the corresponding linear equation from the three linear systems, and by setting $dw_2 = 0$. $g(\bar{w}_1, h^1, Q)$ is therefore identified up to the additive constant and all marginal utilities are identified except u_l^2.

Appendix B: Derivation of Euler Equations when Consumption and Leisure are Strongly Separable

The utility function takes the following form:

$$U(c,l) = \frac{c^{1-\rho}}{1-\rho} + \theta \frac{l^{1-\gamma}}{1-\gamma}.$$

The marginal utilities are therefore,

$$U_c(c,l) = c^{-\rho} \qquad \text{and} \qquad U_l(c,l) = \theta l^{-\gamma}.$$

The function $g(w, T-h)$ that determines consumption can then be written as follows:

$$c = g(w, T-h) = \left(\frac{1}{\theta}\right)^{\rho} w^{\rho} (T-h)^{\frac{\gamma}{\rho}}. \tag{2.30}$$

The standard Euler equation has the following form:

$$\beta E\left[\left(\frac{c_t}{c_{t+1}}\right)^{\rho} R_t\right] = 1.$$

The marginal utility of consumption after replacing c with $g(w, T-h)$ can be written as follows

$$f_c^1 = \frac{\theta^{2\rho}}{w^{2\rho}(T-h)^{\gamma}}.$$

Hence, the transformed Euler equation takes the following form:

$$\beta E\left[\left(\frac{w_t}{w_{t+1}}\right)^{2\rho}\left(\frac{T-h_t}{T-h_{t+1}}\right)^{\gamma} R_t\right] = 1.$$

The intratemporal condition can be derived by taking logs of equation (2.30) to obtain,

$$\ln c = -\rho \ln \theta + \rho \ln w + \frac{\gamma}{\rho}(T-h)$$

We can estimate the standard order equation using the non-linear formulation just derived or the following log-linear approximation:

$$\ln c_{t+1} - \ln c_t = \frac{1}{\rho} \ln \beta + \frac{1}{\rho} \ln R_t + \delta X_t + \epsilon_t.$$

For our approach,

$$\ln (T - h_{t+1}) - \ln (T - h_t) = \frac{1}{\gamma} \ln \beta + \frac{1}{\gamma} \ln R_t - \frac{2\rho}{\gamma} (\ln (w_{t+1}) - \ln (w_t)) + \delta X_t + \epsilon_t.$$

2.11 Tables

Log-linearized Euler Equations for Singles.

Table 2.2: Estimation of Euler Equations for all Households

Ind. Variable	2-stage GMM			2-stage Least Squares		
$\ln(R_{t+1})$	0.148	0.106	0.013	0.156	0.173	0.096
	[0.125]	[0.137]	[0.145]	[0.155]	[0.166]	[0.178]
$\Delta\ln(\text{famsize})$	0.362^{*}	0.367	0.367	0.449^{*}	0.440	0.520
	[0.202]	[0.237]	[0.320]	[0.238]	[0.272]	[0.369]
Δkids	0.043	0.063	0.061	0.006	0.045	0.038
	[0.108]	[0.123]	[0.133]	[0.119]	[0.131]	[0.146]
Δhw	-	-0.552^{**}	-0.239	-	-0.542^{**}	-0.274
		[0.244]	[0.289]		[0.274]	[0.314]
$\Delta\ln(hl)$	-	-	0.249^{*}	-	-	0.333^{**}
			[0.140]			[0.161]
Δww	-	0.278	0.450^{*}	-	0.251	0.384
		[0.231]	[0.265]		[0.233]	[0.277]
$\Delta\ln(wl)$	-	-	-0.037		-	-0.034
			[0.053]			[0.056]
$\ln(y_t)$	0.168^{**}	0.201^{**}	0.202^{**}	0.168^{**}	0.187^{**}	0.206^{**}
	[0.075]	[0.087]	[0.092]	[0.085]	[0.092]	[0.100]
J-Statistic	43.4	36.2	30.7	-	-	-
$P > \chi^2$	0.33	0.55	0.72	-	-	-
n. of observ.	337	337	337	337	337	337
n. of cohort	7	7	7	7	7	7

Asymptotic standard errors are in brackets. All specifications include a constant and three seasonal dummies. The instrument set is the same across columns and includes the first lag of family size growth and of the change in two education dummies, the first equal to one if the head only attended elementary school, the second equal to one if the head attended high school but did not graduate; the first, second, and third lags of nominal municipal bond interest rate, the change in number of children, the change in number of children younger than 2, labor supply growth of the spouse if present, real consumption growth, real municipal bond interest rate, and marginal tax growth; the first, second, third, and fourth lags of the change in dummy equal to one if the head works and in a dummy equal to one if the wife works and is present, nominal 3-month treasury bill rate growth; the second and third lags of salary growth; the second, third and fourth lags of income growth and head's leisure growth. hw and ww are dummies equal to 1 if the head works and if the spouse works. $\ln(hl)$ and $\ln(wl)$ are the logs of head's and spouse's quarterly leisure. y_{t-1}, $y_{h,t-1}$ and $y_{w,t-1}$ are household, head's and spouse's income at $t-1$. (**) and (*) indicate that the coefficient is significant, respectively, at the 5 and 10 percent level.

Table 2.3: Estimation of Euler Equations for For Singles and Couples

Ind. Variable	Singles	Couples	Singles	Couples	Singles	Couples
$\ln(R_{t+1})$	0.029	0.366	0.278	0.596	0.227	0.606
	[0.751]	[0.433]	[0.822]	[0.475]	[0.830]	[0.492]
$\Delta \ln(\text{famsize})$	0.222	0.083	0.152	-0.010	0.278	-0.230
	[0.253]	[0.366]	[0.292]	[0.356]	[0.311]	[0.383]
Δkids	-0.050	0.267^{**}	0.035	0.306^{**}	-0.060	0.368^{**}
	[0.168]	[0.136]	[0.179]	[0.136]	[0.196]	[0.146]
Δhw	-	-	0.976^{**}	0.559^{**}	1.200^{**}	0.354
			[0.378]	[0.291]	[0.421]	[0.332]
$\Delta \ln(hl)$	-	-	-	-	0.457	-0.231
					[0.489]	[0.281]
Δww	-	-	-	0.175	-	0.601^{*}
				[0.189]		[0.347]
$\Delta \ln(wl)$	-	-	-		-	0.708
						[0.514]
$\ln(y_t)$	0.118	0.266^{**}	-0.024	0.187^{**}	-0.025	0.186^{**}
	[0.121]	[0.083]	[0.142]	[0.091]	[0.145]	[0.093]
J-Statistic	32.7	43.5	24.8	38.4	23.9	38.2
$P > \chi^2$	0.53	0.32	0.85	0.45	0.85	0.37
n. of observ.	333	366	333	366	333	366
n. of cohort	7	7	7	7	7	7

Asymptotic standard errors are in brackets. All specifications include a constant and three seasonal dummies. The instrument set is the same across columns and includes the first lag of family size growth and of the change in two education dummies, the first equal to one if the head only attended elementary school, the second equal to one if the head attended high school but did not graduate; the first, second, and third lags of nominal municipal bond interest rate, the change in number of children, the change in number of children younger than 2, labor supply growth of the spouse if present, real consumption growth, real municipal bond interest rate, and marginal tax growth; the first, second, third, and fourth lags of the change in dummy equal to one if the head works and in a dummy equal to one if the wife works and is present, nominal 3-month treasury bill rate growth; the second and third lags of salary growth; the second, third and fourth lags of income growth and head's leisure growth. hw and ww are dummies equal to 1 if the head works and if the spouse works. $\ln(hl)$ and $\ln(wl)$ are the logs of head's and spouse's quarterly leisure. y_{t-1}, $y_{h,t-1}$ and $y_{w,t-1}$ are household, head's and spouse's income at $t-1$. (**) and (*) indicate that the coefficient is significant, respectively, at the 5 and 10 percent level.

Table 2.4: Log-linearized Individual Euler Equations for Single Females: Identification Result vs Standard Methods.

Parameters	Identification	Standard
ρ	5.02	5.21
	[2.49]	[1.08]
σ	0.30	0.12
	[0.44]	[0.07]
J-Statistics	15.2	37.9
$P > \chi^2$	0.71	0.73
number of observations		1228

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first to second lags of after tax real wage growth, marginal tax growth; first to fourth lags of real consumption growth, income growth, gross pay growth, labor supply growth, the household specific price index growth. All instruments are calculated at the cohort level.

Table 2.5: Log-linearized Individual Euler Equations for Single Males: Identification Result vs Standard Methods.

Parameters	Identification	Standard
ρ	1.72	1.65
	[0.96]	[0.43]
σ	0.08	0.55
	[0.73]	[0.26]
J-Statistics	9.5	33.8
$P > \chi^2$	0.96	0.87
number of observations		1138

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first to second lags of after tax real wage growth, marginal tax growth; first to fourth lags of real consumption growth, income growth, gross pay growth, labor supply growth, the household specific price index growth. All instruments are calculated at the cohort level.

Non-linear Euler Equations for Singles

Table 2.6: Individual Euler Equations for Single Females: Identification Result vs Standard Methods.

Parameters	Identification	Standard
ρ	5.26	5.44
	[0.57]	[0.72]
σ	0.08	0.31
	[0.12]	[0.06]
J-Statistics	14.6	34.7
$P > \chi^2$	0.33	0.34
number of observations		1228

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first lag of after tax real wage growth; first to third lags of real consumption growth, marginal tax growth, real gross interest rate growth; first to fourth lags of income growth, the household specific price index growth; all instruments are calculated at the cohort level.

Table 2.7: Individual Euler Equations for Single Males: Identification Result vs Standard Methods.

Parameters	Identification	Standard
ρ	1.96	1.62
	[0.63]	[0.39]
σ	0.45	0.51
	[0.56]	[0.24]
J-Statistics	5.5	20.9
$P > \chi^2$	0.85	0.75
number of observations		1138

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first and second lags of labor supply growth; first to third lags of leisure growth; first to fourth lags of income growth, log of real gross rate of return; third and fourth lags of real consumption growth; all instruments are calculated at the cohort level.

Log-linearized Euler Equations for Singles, Controlling for Selection.

Table 2.8: Log-linearized Individual Euler Equations for Single Females: Identification Result vs Standard Methods Controlling for Selection.

Parameters	Identification	Standard
ρ	5.40	5.03
	[2.27]	[1.23]
σ	0.16	0.10
	[0.15]	[0.05]
Inverse Mills' Ratio t	2.12	4.24
	[2.10]	[3.45]
Inverse Mills' Ratio $t+1$	-1.69	-4.47
	[5.22]	[3.93]
J-Statistics	14.22	33.5
$P > \chi^2$	0.58	0.79
number of observations		1228

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first and second lags of labor supply growth; first to third lags of leisure growth; first to fourth lags of income growth, log of real gross rate of return; third and fourth lags of real consumption growth; all instruments are calculated at the cohort level.

Table 2.9: Log-linearized Individual Euler Equations for Single Males: Identification Result vs Standard Methods Controlling for Selection.

Parameters	Identification	Standard
ρ	2.14	1.70
	[1.09]	[0.31]
σ	0.10	0.18
	[0.20]	[0.09]
Inverse Mills' Ratio t	0.19	-0.53
	[0.75]	[0.88]
Inverse Mills' Ratio $t+1$	-0.65	4.62
	[2.14]	[3.76]
J-Statistics	9.3	33.2
$P > \chi^2$	0.90	0.80
number of observations		1138

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first and second lags of labor supply growth; first to third lags of leisure growth; first to fourth lags of income growth, log of real gross rate of return; third and fourth lags of real consumption growth; all instruments are calculated at the cohort level.

Non-linear Euler Equations for Singles, Controlling for Selection.

Table 2.10: Non-linear Individual Euler Equations for Single Females: Identification Result vs Standard Methods Controlling for Selection.

Parameters	Identification	Standard
ρ	5.41	5.13
	[0.85]	[1.11]
σ	0.14	0.32
	[0.19]	[0.08]
Inverse Mills' Ratio t	-0.74	-0.50
	[4.09]	[2.34]
Inverse Mills' Ratio $t+1$	1.29	2.33
	[6.38]	[4.36]
J-Statistics	14.9	32.6
$P > \chi^2$	0.14	0.29
number of observations		1228

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first and second lags of labor supply growth; first to third lags of leisure growth; first to fourth lags of income growth, log of real gross rate of return; third and fourth lags of real consumption growth; all instruments are calculated at the cohort level.

Table 2.11: Non-linear Individual Euler Equations for Single Males: Identification Result vs Standard Methods Controlling for Selection.

Parameters	Identification	Standard
ρ	1.90	1.84
	[0.65]	[0.29]
σ	0.45	0.26
	[0.35]	[0.15]
Inverse Mills' Ratio t	-0.40	2.08
	[2.63]	[2.65]
Inverse Mills' Ratio $t+1$	0.47	-1.84
	[1.76]	[2.10]
J-Statistics	5.62	23.4
$P > \chi^2$	0.59	0.44
number of observations		1138

Asymptotic standard errors in brackets. All models are estimated with GMM using the following instruments: first and second lags of labor supply growth; first to third lags of leisure growth; first to fourth lags of income growth, log of real gross rate of return; third and fourth lags of real consumption growth; all instruments are calculated at the cohort level.

Euler Equations for Couples.

Table 2.12: Individual Euler Equations for Couples with the Efficiency Condition.

Parameters	Wife	Husband	Parameter Difference	Unitary Model with Separability
ρ	4.42	2.51	1.91	3.69
	[0.43]	[0.78]	[0.88]	[0.40]
σ	0.29	0.27	0.02	-
	[0.12]	[0.13]	[0.19]	
μ	0.82	-	-	-
	[4.32]			
J-Statistics			56.9	24.4
$P > \chi^2$			0.33	0.55
number of observations				1266

Asymptotic standard errors in brackets. The estimate of μ is obtained by computing a first order Taylor expansion under the assumption that the constants in the measurements errors are equal to zero. All models are estimated with GMM using the following instruments: first to second lags of marginal tax growth; first to third lags of wife's and husband's gross pay growth; first to fourth lags of real household consumption growth, household income growth, wife's and husband's after tax real wage growth, wife's and husband's labor supply growth.

Table 2.13: Individual Euler Equations for Couples without the Efficiency Condition.

Parameters	Wife	Husband	Parameter Difference
ρ	4.23	2.62	1.61
	[0.43]	[0.66]	[0.78]
σ	0.24	0.17	0.07
	[0.13]	[0.13]	[0.18]
μ	-	-	-
J-Statistics			56.1
$P > \chi^2$			0.36
Efficiency Test			
Distance Statistics			0.8
$P > \chi^2$			0.37
number of observations			1266

Asymptotic standard errors in brackets. The estimate of μ is obtained by computing a first order Taylor expansion under the assumption that the constants in the measurements errors are equal to zero. All models are estimated with GMM using the following instruments: first to second lags of marginal tax growth; first to third lags of wife's and husband's gross pay growth; first to fourth lags of real household consumption growth, household income growth, wife's and husband's after tax real wage growth, wife's and husband's labor supply growth.

CHAPTER 3

The Effects of Paternity Leave: Evidence from the Introduction of a Father's Quota in Quebec

3.1 Introduction

In this paper, I analyze early-life childcare decisions within the household. Understanding how households make these decisions is an important consideration when analyzing labor market outcomes of parents. There is considerable evidence that temporary reductions in labor supply, in the form of parental leave for instance, have a major impact on earnings and careers. Indeed, studies argue that mothers, that take most parental leave, pay large child earnings penalties that persist over the long-term (see for example Kleven et al. (2019)). In this context, the main question this paper tries to answer is how households make decisions related to parental leave. More precisely, this paper will study the effect of incentivizing fathers to take more parental leave. It will also analyze how parental leave decisions are related to the labor market income of each parent and if there is evidence of specialization within households.

To answer these questions, I will use a major change in parental leave policy in Quebec. In January 2006, Quebec's provincial government created the "Régime québécois d'assurance parentale" or Quebec Parental Insurance Plan (QPIP). One of the main components of the reform was the introduction of a "father's quota": 5 weeks of non-transferable paid parental leave weeks reserved for fathers. Prior to the reform, parental leave time could be split between the parents without any restrictions. However, in practice, in the large majority of households, only mothers took paid parental leave. The father's quota incentivizes fathers to take time off from work at the birth of a child. Since the weeks are not transferable to mothers, not taking them would result in the household foregoing some of the available paid parental leave time. Indeed, the policy had a significant impact on the household's decisions, as the fraction of fathers taking some parental leave time more than doubled following the reform. The 2006 reform provides the ideal setting to analyze household parental leave decisions. First of all, analyzing the response of fathers to the new incentives introduced by the father's quota provides insight into how households make decisions related to splitting parental leave time. Second, this paper, with the help of detailed administrative data on QPIP recipients, is able to provide evidence on how parental leave decisions are related to parents' income.

This paper makes the following main contributions. First, it provides evidence that while the reform changed many aspects of the parental leave system the most important aspect is the introduction of the father's quota. It shows that while the reform had an impact on both fathers and mothers, the magnitude of the impact was much larger for fathers than for mothers. Using a difference-in-difference estimator, the reform is estimated to induce an extra 50% of all new fathers to take parental leave. After 2006 in Quebec, more than 80% of all new fathers take some parental leave compared to 30% beforehand. Meanwhile, the fraction of mothers taking some parental leave goes from around 85% to around 95%. This paper then examines parental leave behavior under QPIP using an administrative dataset. It finds that the majority of households split the weeks of parental leave in a very specific way: the mother takes the maximum amount available to her and then the father takes the weeks that remain. This behavior seems to imply that the constraint introduced by the father's quota binds and that most households would prefer to specialize even further. The paper then analyzes the relationship between income and parental leave behavior. While taking parental leave is more costly for higher income households (in terms of foregone income), these households still take more parental leave. While richer household take more parental leave overall (summing mother and father time-off), when focusing on fathers this pattern does not hold. In fact, fathers that have higher income take less time off. This suggests a higher degree of specialization in households with high income fathers.

Understanding parental leave decisions is crucial to analyze labor market outcomes of parents. A major question is whether a policy such as a father's quota can reduce the wage-gap between men and women. There exists a large literature that tries to determine the causes of the wage-gap. Blau and Kahn (2017) provide an overview of the literature that analyzes its determinants. One explanation that is particularly compelling is provided in Goldin (2014). The author of that paper argues that the main explanation for the remaining gender wage-gap is the remuneration structure of the labor market. Firms disproportionately reward workers that work long and unusual hours. As a result, the penalty of having a flexible job, in terms of foregone wages, is large. For a variety of reasons, women might have a preference, relative to men, for flexible schedules. Therefore, Goldin argues that women are hit disproportionately by these penalties for flexible schedules or time off.

Goldin's paper provides a framework to think about the gender wage-gap. However, it does not address the question of why women value flexibility more than men. One possible explanation is that women have to take time off to have children and that early-life childcare responsibilities fall mainly on mothers. Therefore, women have a harder time than men conciliating family and professional life and end up paying penalties in terms of lower wages. Bronson (2015), provides evidence that this is an important consideration for many women, as it appears to be a major mechanism in explaining differences in college major choices between men and women. In fact, Bronson's paper argues that women select disproportionately into degrees that are lower paying but allow a high degree of work-family flexibility. In addition to affecting education decisions, many studies find evidence of a "child earnings penalty", a decrease in earnings growth following the birth of the first child. For example, Waldfogel (1997) finds that even after controlling for actual labor market experience, such a penalty can still be observed. More recently Kleven et al. (2019) estimate that in Denmark the "child penalty" is about 20% of earnings and accounts for about 80% of total earning inequalities. Finally, Gruber (1994) shows that the cost of mandated benefits related to childbirth where largely passed on to mothers in the form of lower wages. Therefore, it is a legitimate question to wonder if the same phenomenon takes place with the father's quota.

Given these observations policies surrounding pregnancy and childcare, such as the father's quota, could have a major impact on the wage-gap. Previous papers have studied the father's quota in other countries. In addition to Quebec, it has been instituted in places such as Sweden, Norway and Iceland. Ekberg et al. (2013) find that in Sweden, the policy had strong effects in the short-term in terms of the splitting of parental leave time. However, they find little evidence of long term effects in terms of division of household work or labor market outcomes. Similarly, looking at Norway, Cools et al. (2015) find little evidence of a change in the division of household work and surprisingly find a negative impact on mothers' labor market outcomes. One possible explanation though is that the reform Cools et al. look at also increased mothers' parental leave time making the effect they identify unclear. Dahl et al. also look at the reform in Norway. They show the presence of important peer effects in the take-up of parental leave for fathers: fathers that know other fathers that took paternity leave are more likely to take some themselves. Finally, Patnaik (2019) looked at the daddy quota policy in Quebec. Her paper shows a strong effect in terms of participation in the program and also a long-term effect in terms of splitting household work.

This paper expands on Patnaik in terms of looking more closely at participation in the program. Patnaik's small dataset (about 200 observations per year) does not allow her to study the determinants of participation in the program. By using an administrative dataset that tracks all program participants I am able to look more closely at how participants self-select by choosing the number of weeks of parental leave they take. One potential caveat is that, given results in previous studies, it is possible that the effects on labor market outcomes are small or insignificant. However, it is still worth looking into.

Finally, a recent report by Clavet, Corneau-Tremblay, and Lacroix (2016) uses the same administrative dataset used here. They conduct a comprehensive review of the effects of the 2006 parental leave reform. One new result presented in this paper is that, conditional on household income, fathers with higher income take less weeks of parental leave on average.

The rest of the chapter is organized as follows. Section 2 describes in detail parental leave systems in Quebec before and after the introduction of QPIP. Section 3 discusses the data that is used in this paper. Section 4 looks at the effect of the reform on parental leave take-up of fathers and mothers. Section 5 describes in detail the behavior of households under QPIP.

3.2 Description of Parental Leave in Quebec

As explained previously, in January 2006, Quebec's provincial government introduced a major reform of the parental leave system. Up until 2006, Quebec's parental leave policy, like that of other Canadian provinces, was under the federal government's Employment Insurance (EI) system. In 2006, Quebec's provincial government created a new system called QPIP. In addition to introducing a father's quota, the new system also changed parental leave policy along many dimensions. These changes are summarized in Table 3.1. It is also useful to note that to this day, other Canadian provinces are still under the EI system. This provides a useful comparison group to think about the effects of the reform.

Table 3.1: Description of Policy Changes

Policy feature	**EI system (pre-2006)**	**QPIP basic regime (post-2006)**	**QPIP special regime (post-2006)**
Total number of weeks	50 weeks	55 weeks	43 weeks
Income replacement rate	55%	70% for 30 weeks then 55% for 25 weeks	75%
Insurable income max (2006)	39,000 C$	57,000 C$	57,000 C$
Father's quota	None	5 weeks	3 weeks

The first thing to notice is that under the new QPIP system, households have the choice between two regimes: a basic regime and a special regime. The trade-off between the two regimes is that the basic regime has more paid parental leave weeks but the income replacement rate is higher in the special regime. Most households choose the basic regime (74%). Households can only choose one regime for both parents and cannot switch in the middle of the parental leave period. Table 3.1 also shows that parental leave policy became more generous following the reform. The maximum number of weeks of paid parental leave increased post 2006 (55 weeks in the basic regime). In addition, total transfers also increased with both the income replacement rate and the insurable income maximum higher under QPIP. The insurable income maximum refers to the maximum income that can be replaced at the specified income replacement rates. Income above this limit is not replaced when parents take parental leave (the replacement rate falls to 0%). The insurable income maximum has increased steadily in both systems since 2006, with QPIP staying more generous. In 2016 they were 51,300 C$ and 71,500 C$ for EI and QPIP respectively. Finally, the weeks reserved for fathers are 5 weeks and 3 weeks under the basic and the special regime respectively.

The fact that parental leave system changed in many ways presents challenges in terms of identifying the effect of the father's quota, since under the more generous plan both mothers and fathers are expected to take more parental leave. In addition, it becomes unclear if fathers take more parental leave because of the father's quota or because of the more generous conditions. However, as shown in Figure 3.1, the policy had a much bigger impact on fathers than mothers.[1] Therefore, while it is important to think about this carefully in the analysis, the main effect of this policy is an increase in paternity leave relative to maternity leave. In any case, the next sections will provide a more detailed description of household decision-making under QPIP.

3.3 Data

This paper uses two main data sources to study parental leave in Quebec. The first source is an administrative dataset that contains the universe of QPIP recipients and detailed information on number of weeks and transfer amounts for program participants. As far as I know, this paper is the first to use this data. The second source is survey data from Statistics Canada. The survey is the Employment Insurance Coverage Survey (EICS). EICS is a supplement to the Labor Force Survey (the main labor force survey

[1]Note that in Figure 3.1, the year refers to the year in which the parents took the parental leave. However, the new parental leave system only applied to children born after January 1st. As a result, the data for 2006 includes both pre and post reform families, which explains why the fraction of parents that take the parental leave is somewhat in between pre and post reform fractions.

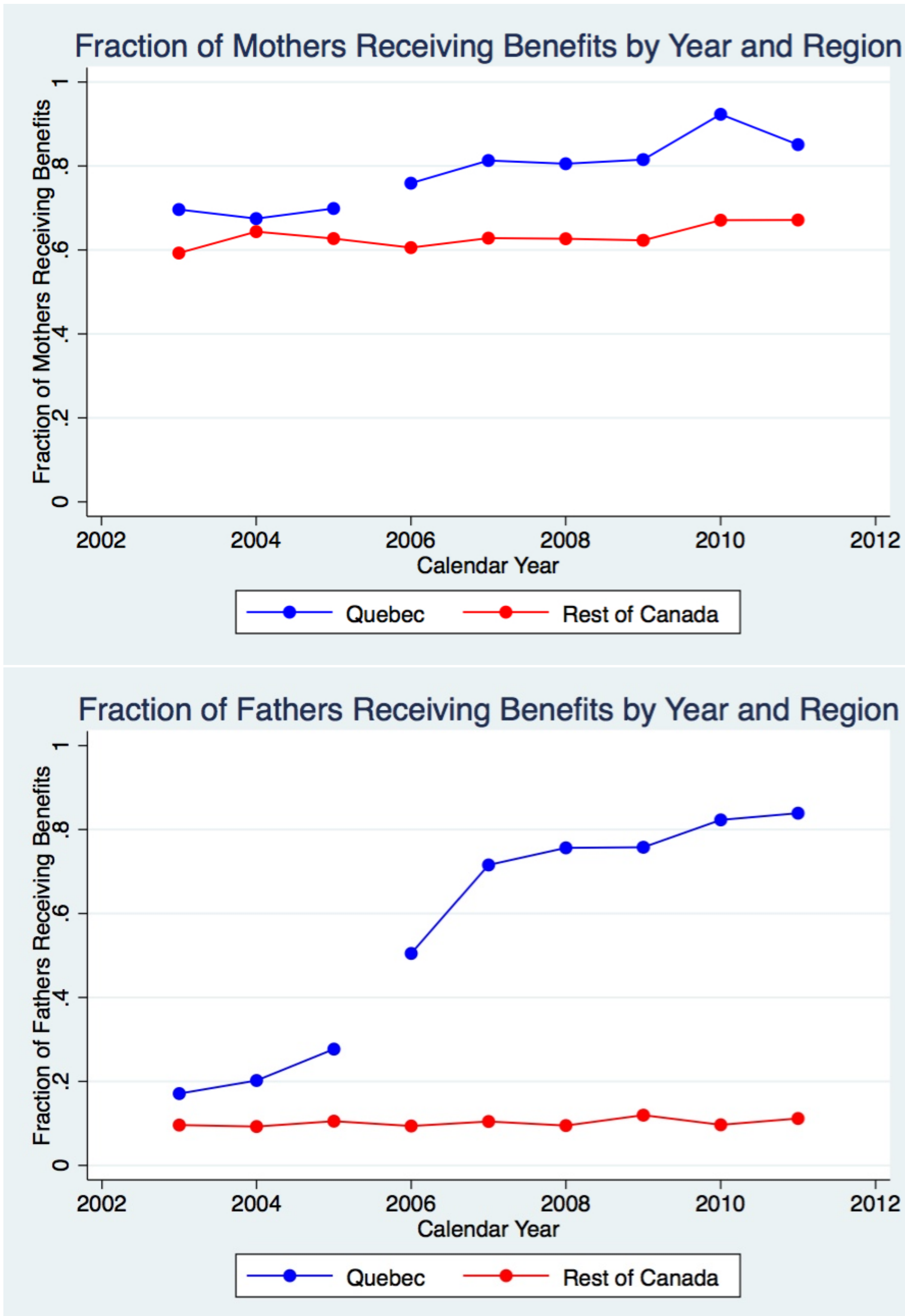


Figure 3.1: Fraction of Parents Receiving Benefits by Year and Region
Data from Employment Insurance Coverage Survey (EICS)

in Canada) that is designed to study participants in cash transfer programs such as parental leave programs.

I will now describe in greater detail each of the two main datasets:

- The administrative data comes from the "Conseil de gestion de l'assurance parentale" (CGAP) that manages QPIP. This dataset allows me to look in detail at early childcare decisions within the household, after the reform. It contains all beneficiaries from QPIP between the introduction of the new system in 2006 and 2015, with information about the number of weeks of parental leave (with the income replacement rate of each week taken), the amount of benefits received and insurable income. It also contains the age of the parent and whether the parent is salaried or self-employed. This dataset allows me to study how households split parental leave under the program and how a household's income might affect decisions.

  However, the data does have some limitations that need to be addressed. The first drawback is that parents that do not take parental leave are not observed in the sample. This means that many observations contain only one parent. Out of 740,000 total births, for 75,000 observations (10.1%) we only observe the father and for 175,000(23.6%) observations we only observe the mother. In the cases where only the father is observed it is probably fair to assume that most cases correspond to households where the mother is ineligible for parental leave because she is either unemployed or not in the labor force. This is supported by the fact that survey data indicates that the participation rate of eligible mothers in the program is higher than 95%. The cases where only mothers are observed is trickier as survey data indicates that around 5% are single mothers and 8% correspond to cases where fathers are not eligible. However, even after the reform, the participation rate of eligible fathers is still only around 85% meaning that there is a significant amount of fathers that make the decision not to participate in the program. Using a back of the envelope calculation (85% of the $100\% - 8\% - 5\% = 87\%$ of eligible fathers) indicates that probably about 13% of the sample are made up of fathers that decide not to participate in the program. The uncertainty around the reason why the father is not observed is a drawback. In particular, it is reasonable to think that all three groups are very different. For example, single mother households probably do not behave in the same way as households where the father decides not to take parental leave. However, even limiting our analysis to the cases where the father is observed provides interesting insights.

  A second limitation is that actual earnings are not observed. Rather only insurable income is. In most cases, this is a good measure of labor market income. However, this is only true up to the maximum insurable income. For people earning more, the maximum insurable income value is reported. Over the course of the period considered, this concerns about 9.6% of observed mothers and 20.4% of observed fathers. The percentages are stable from year to year. This is a problem when considering the effects of income on parental leave decision. However, the very large dataset makes it possible to consider subsamples of the population that are not affected by this censoring problem.

- The Employment Insurance Coverage Survey (EICS) has been active since 1997 and is linked to the broader Labor Force Survey. It gives detailed information on a sample of mothers receiving parental leave benefits, including number of weeks they received benefits, amount of benefits, and pre-leave wage. It also says whether the spouse is eligible for parental leave, and if he is, whether he took some leave. In addition, it contains demographic information about the household including province of residence. The survey covers all Canadian provinces before and after 2006, allowing cross-province comparisons that are a good starting point to look at the effect of the reform (as seen in figures 1 and 2 for example). The data does have some limitations: since it is not designed to look specifically at the program in Quebec, the sample of new parents in Quebec is only about 250 per year, which might be small if we worry that there might be a lot of heterogeneity in parental leave decisions. Also, as stated previously, 2006 is problematic as the survey year refers to the year in which benefits were received and not year of birth. As a result 2006 contains some households that are under the EI system and some that are under the QPIP system. The way I will deal with this is by dropping 2006 in certain cases. Finally, the information on spouse behavior is rather limited as we only know whether or not the spouse took parental leave and do not for how long or the amount of program transfers. This limits the amount of analysis that can be done in terms of understanding the full extent of household decision-making. Nevertheless, it is a start and provides some evidence that complements other data sources.

3.4 Effect of the Reform on Parental Leave Take-up

As stated previously the introduction of the reform in Quebec but not in other Canadian provinces suggests thinking of the reform as a treatment with Quebec being the treated group and other Canadian provinces as the control group. Therefore, in general, the empirical strategy will rely on comparing the change in parental leave behavior of parents that had children after January 1st 2006 in Quebec to the change in parental leave behavior in other Canadian provinces.

For such a quasi-experimental setting to identify the average treatment effect requires identifying assumptions. Crucially, it relies on the fact that households cannot influence which group they fall in. Since date of birth is hard to control this is a reasonable assumption. Migration is another phenomenon that might confound the estimated effect. However, migration between Quebec and other Canadian provinces is not very large, with the total number of interprovincial migrants (both in and out of Quebec) less than 0.5% of the population. In addition, the new parental leave system is not so much more generous that we would expect parents to delay having children or migrate in response to it. Finally, looking at the summary statistics there is also almost no change in the population in terms of wages, age, education or size of family in the year before and the year after the reform. Taken together these elements suggest that the quasi-experimental setting identifies the average treatment effect.

Figure 3.1 suggests that there was a major impact of the reform in terms of take-up of parental leave, particularly for fathers. The fraction of fathers taking some parental leave in Quebec increases from less than 30% before 2006 to close to 80% from 2007 onwards. There is also an effect on mothers, as the fraction taking some parental leave in Quebec increases from 70% to more than 80%. In contrast, in other Canadian provinces, the fraction of parents taking parental leave is very stable over the whole period. Figure 3.2 presents similar graphics for the subpopulation of parents that are eligible for parental leave. The main reason a new parents would not be eligible for parental leave is if they earned less than 2000 C$ in wages in the 52 weeks preceding the start of the leave period. Therefore, this concerns mostly parents that are not in the labor force or are long-term unemployed. The evidence in Figure 3.2 is similar to that in figures Figure 3.1: there is an impact on take-up for parents with a much bigger impact for fathers. The fraction of eligible fathers that take some parental leave increases from close to 30% to more than 80% (and more than 90% by 2011). For mothers, the fraction increases from 90% to almost 100%.

A more formal approach would be to use a difference-in-difference method to estimate the impact of the reform. A difference-in-difference estimate can be obtained by estimating the following equation using the EICS data separately for fathers and mothers:

$$p_i = \delta(Q_i \times \text{Post-treatment}_i) + \lambda Q_i + \sum_T \gamma_T Y_i + X_i'\beta + \epsilon_i \tag{3.1}$$

p_i is the dependent variable in the specification. In two specifications it will be an indicator variable equal to 1 if parent i (either the mother or the father) takes some parental leave. In the third specification p_i refers to the number of months of parental leave that mother i takes. Q_i is an indicator variable equal to 1 if parent i lives in

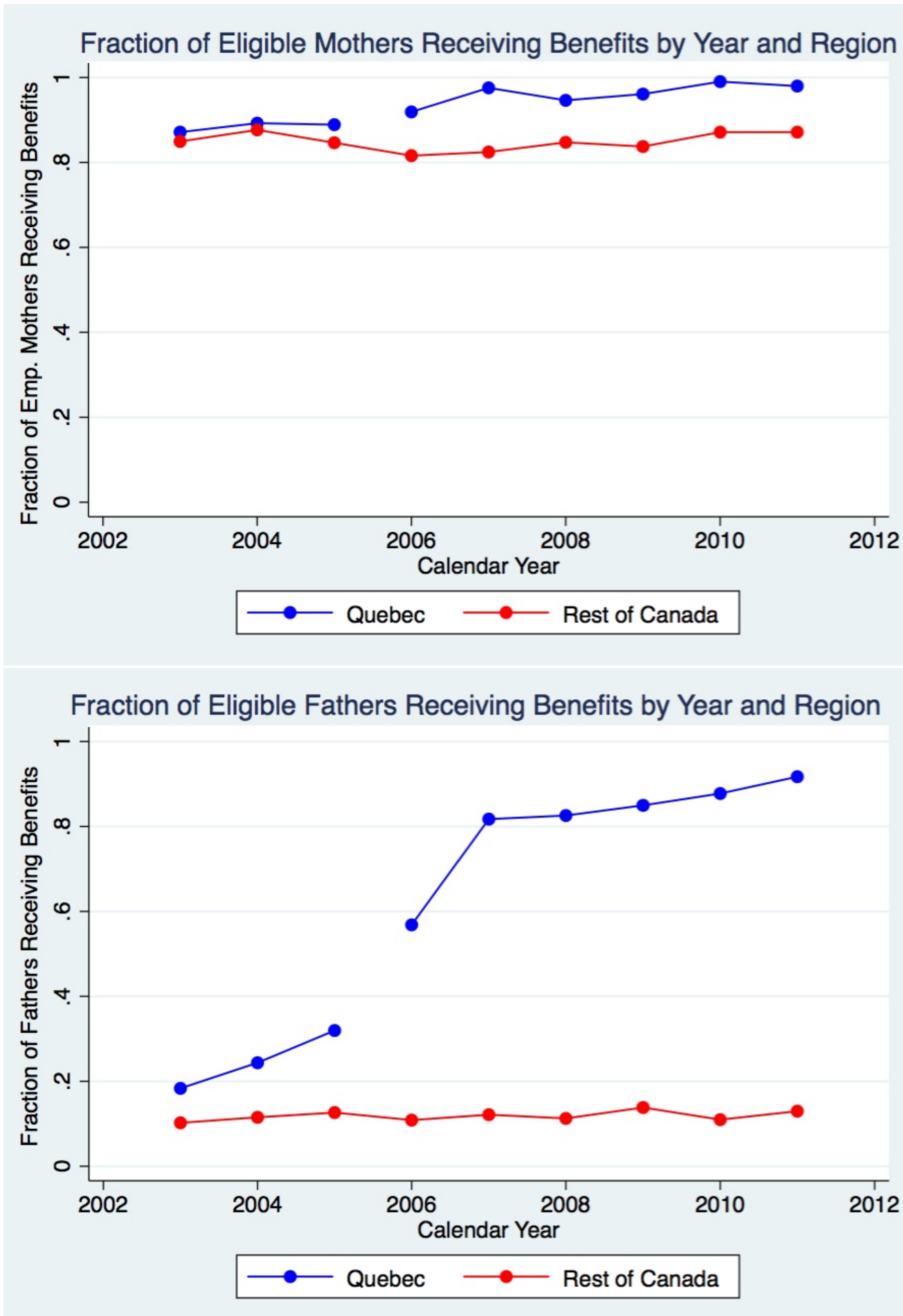


Figure 3.2: Fraction of Eligible Parents Receiving Benefits by Year and Region
Data from Employment Insurance Coverage Survey (EICS)

Quebec. Post-Treatment$_i$ is an indicator variable equal to 1 if parent i was eligible for parental leave in 2007 or later (after the introduction of QPIP). Therefore, the interaction of Q_i and Post-Treatment$_i$ will be equal to 1 only for our treated group (Quebec residents after the introduction of QPIP). Y_i are year-specific indicator variables. X_i is a vector of demographic controls that include (in some specifications) hourly wage, education controls and household total income[2]. This estimation excludes data from 2006 as the survey does not distinguish parents under QPIP and under EI. The estimated difference-in-difference effect of the reform is then the estimate for δ. The results from the estimation are presented in Table 3.2.

Table 3.2: Effect of the Reform on Parental Leave Participation

	Father	Mother	Mother months	Father	Mother	Mother months
QC × Post-2006	0.554***	0.0941***	0.126	0.568***	0.101***	-0.0981
	(0.0226)	(0.0179)	(0.267)	(0.0247)	(0.0182)	(0.265)
Quebec	0.119***	0.0272	0.327	0.124***	0.0221	0.231
	(0.0181)	(0.0160)	(0.226)	(0.0202)	(0.0162)	(0.229)
Hourly Earning				0.000867	0.00234***	-0.00170
				(0.000495)	(0.000552)	(0.00672)
Family Size				-0.00880	-0.0258***	-0.274***
				(0.00523)	(0.00556)	(0.0721)
Income group				0.00626	0.0453***	0.274***
				(0.00460)	(0.00485)	(0.0625)
Constant	0.0876***	0.848***	9.834***	0.0812**	0.651***	10.36***
	(0.0100)	(0.0125)	(0.173)	(0.0304)	(0.0332)	(0.423)
Observations	8284	7292	6551	6761	6747	5698
R^2	0.321	0.017	0.005	0.354	0.069	0.012

Heteroskedasticity robust standard errors in parentheses, also includes year fixed effects and education controls
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As seen in Table 3.2, the estimates do not depend on the specification. The estimated increase in the probability that fathers take some parental leave is close to 0.55. The same number for mothers is close to 0.10. For mothers, richer households take more time off, although the effect is economically small (a 40 C$ increase in hourly wage - 4 times the standard deviation - is only associated with approximately a 3 day increase in parental leave). Finally, this specification does not provide any evidence that the reform changed the amount of months of parental leave that mothers took. This could be because parental leave time is not measured very precisely (at the month level). However, it is also consistent with the fact that QPIP gave a maximum of 50 weeks to the mother (under the basic regime, which most households choose) which is the same as the total number of weeks under EI.

[2]Total household income is grouped in 20,000/year increments. This is the most precise measurement of household income that is available in this data

One potential issue with the method described above can be seen in Figure 3.1 and Figure 3.2. The identifying assumption of the difference-in-difference estimator is that without the reform, the treatment and control group would have evolved following parallel trends. Looking at those figures, it seems like Quebec had been experiencing an increase in the fraction of fathers taking parental leave relative to the rest of Canada even before 2006. However, I perform a variety of tests on whether the pre-treatment trend is significant or not and I find that they all fail to reject the null hypothesis of no difference in trends. The first test I perform is simply comparing the proportion of fathers in Quebec that take parental leave in 2004 and 2005, with the null hypothesis that the proportions are equal. I find p-values of 0.087 and 0.129 for the total population of fathers and the subpopulation of eligible fathers respectively. The difference between the two proportions is therefore not significant at the 5% level. Next, I estimate the linear trend for the pre-2006 data on father's leave. I use data going back to 2000 to estimate the trend more precisely.[3] I allow the trend to be different for Quebec and for other Canadian provinces and then test whether the trends are statistically different or not. I find p-values of 0.059 and 0.094 for the total population of fathers and the subpopulation of eligible fathers respectively. Again, the null hypothesis that the two pre-treatment trends are the same cannot be rejected.

The tests suggest that the difference-in-difference approach is adequate. However, the fact that the p-values are so close to the thresholds might still be worrying. One robustness check that I perform is attempting to construct a synthetic control (Abadie et al. 2010) based on a weighted average of other provinces. However, in this particular case, this method does not work because no other province experienced a growth similar to Quebec's in terms of increased paternity leave. In this respect Quebec is an outlier. Therefore, there is no way to reweigh the provinces to get an adequate synthetic control. Given this issue, the most that can be said is that their is no statistical evidence that the pre-treatment trends are different in Quebec and in other Canadian provinces.

3.5 Parental Leave Behavior under QPIP

The administrative dataset from CGAP allows us to study in more detail the way in which different households take parental leave. In particular, it allows us to break down the way in which weeks are shared between parents and the relationship between earnings and parental leave decisions.

[3] I also perform the same tests with data only between 2003 and 2005 and find almost identical results - all fail to reject the null of no difference in the trends at the 5% level, although barely.

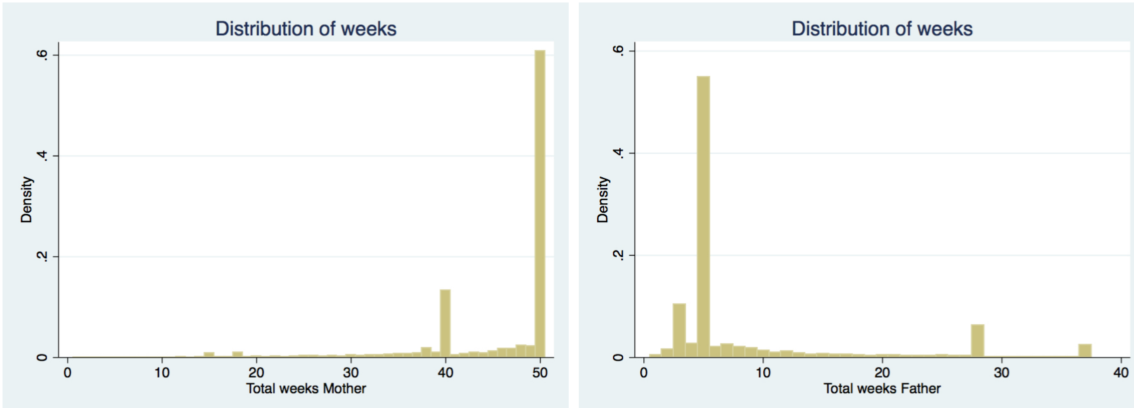


Figure 3.3: Histogram of Parental Leave Weeks

3.5.1 Sharing of Parental Leave within the Household

Figure 3.3 presents a histogram of the distribution of weeks for fathers and mothers. A lot of information can be gathered from this simple graph. First, 60.8% of mothers take 50 weeks of parental leave - the maximal amount under the basic regime. Another 13.2% take 40 weeks, the maximal amount of weeks under the special regime. 13.3% take lower than 40 weeks with no particular value standing out. Similarly 12.7% take an amount between 40 and 50 weeks with most of the mass concentrated closer to 50 weeks.[4] Fathers also have a few values that stand out: 55% of fathers take 5 weeks and 10.4% take 3 weeks. These are the maximal number of weeks that fathers can take if the mother is taking her maximum in the basic and special regimes respectively. However, two other values stand out for fathers: 28 weeks (6.3%) and 37 weeks (2.5%). These are the maximum amount of weeks that fathers can take under the special and basic regimes respectively.[5] Interestingly, for 79.5% of the households where fathers take either 28 or 37 weeks the mother is not observed. Given the participation rate of eligible mothers close to 100% it is safe to assume that in most of these cases the mother is not eligible. Finally, most of the remaining mass of the distribution is concentrated between 3 and 15 weeks. Figure 3.4 and Figure 3.5 present histograms broken down by type of regime and tell a similar story.

Given all these observations, a lot can be said about the way most households make decisions. Although most households choose the basic regime, a substantial portion

[4]This is consistent with maximizing behavior since by taking 41 weeks the household has to choose the basic regime which has a lower income replacement rate over all weeks. Therefore, the marginal cost of this extra week is very high.

[5]15 and 18 weeks are reserved for mothers as pregnancy leave under the special and basic regimes respectively

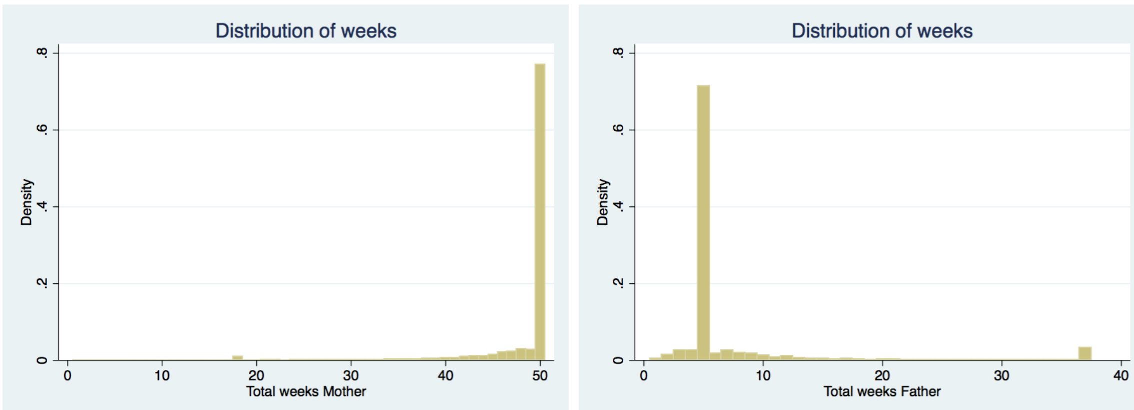


Figure 3.4: Histogram of Parental Leave Weeks Under Basic Regime

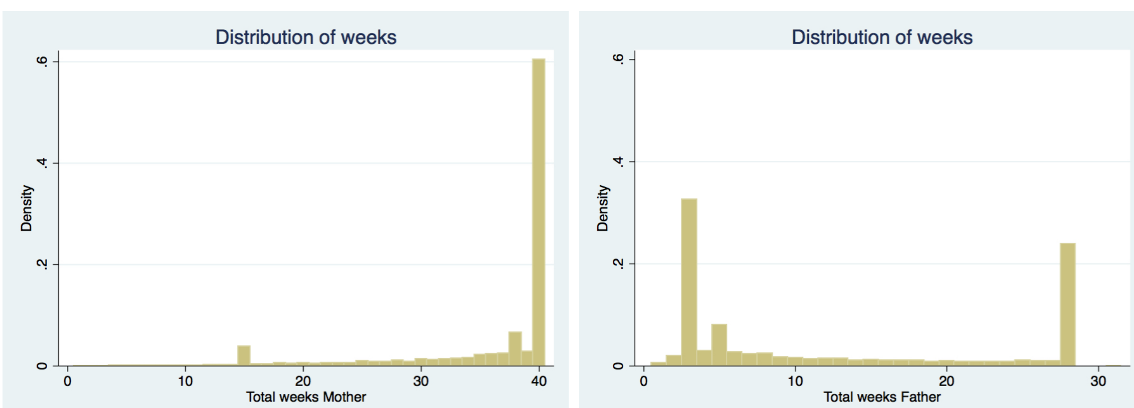


Figure 3.5: Histogram of Parental Leave Weeks Under Special Regime

choose the special regime. A large fraction of households take the maximal amount of leave (or very close to the maximal amount at least). The way most households split the weeks is by allocating all the parental leave that can be freely shared to mothers and then fathers taking the maximum amount of weeks that cannot be shared. This suggests that the constraint imposed by the father's quota is binding and that households would rather specialize fully when it comes to parental leave. This is also suggestive evidence that some of the political motivations for this policy might not be working out. For example, one could think that such a policy would remove a fixed cost of taking any parental leave for fathers (in the form of stigma from employers for instance) and allow households to better optimize the sharing of parental leave. However, the fact that a large proportion of fathers are taking some leave but few are taking more than the amount reserved for them does not support that idea.

3.5.2 Parental Leave and Insurable Income

Another interesting dimension of household decision-making is: how does length of parental leave vary with insurable income? In particular, it is interesting to see if there are observable patterns in terms of how benefit take up is related to earnings. One important related question is the distributive effect of the policy. Dahl et al. (2013) argue that in Norway paid maternity leave is regressive with transfers benefiting mostly higher income households. QPIP is funded by a payroll tax that is proportional to insurable income. Since transfers are also proportional to insurable income, the question of the distributive effect comes down in large part to the number of weeks taken by the household as a function of incomes.

As most households take the maximum amount of parental leave - or close to that amount - one way in which households vary the amount of weeks they take is by choosing either the basic regime or the special regime. Table 3.3 shows that average insurable income is higher for households that choose the basic regime. However, one thing to keep in mind is that in many ways the two groups might not be comparable. In particular, many households that choose the special program are probably single income households (single mothers or mothers not in the labor force). This can be seen by the fact that for 42% of households that choose the special regime only one parent is observed, In comparison, only 18.5% of households that choose the basic regime have only one observable parent. Some of those cases might be fathers that decide not to take any parental leave. However, the high participation rate observed in the survey data implies that it is not the majority. Therefore, this provides evidence that the special regime is used by single income families to take leave without foregoing too much income. Another way in which this can be seen is that 74.5% of fathers choose the special regime when the mother is not observed (as argued previously, these are mostly cases where the mother is not in the labor force).

Table 3.3: Mean Insurable Income by Regime Type

	Basic	Special
Father's Weekly Insurable Income	926	847
Mother's Weekly Insurable Income	730	614

Another way to approach this question is to look more directly at the relationship between number of weeks of parental leave and insurable income. One issue is that households with different structures - such as single mothers - or with different labor force status are not eligible for the same number of weeks. To get around this, Table 3.4 presents the regression of total number of weeks on insurable incomes and restricts the analysis to the subset of the population that choose the basic regime and where both parents are observed. This population is relatively comparable and also has the advantage that both parents' incomes are observed. Interestingly, higher weekly insurable income for mothers is related to more total parental leave, while higher insurable income for fathers is associated with less parental leave. Quantitatively the effect for mother is much more important. Taken together, these results suggest that overall richer households take more parental leave time. This is true even though higher earning households pay a higher cost for parental leave in terms of non-replaced income. The results also imply that for a given total household income, the father earning more implies a decrease in total weeks. This could be explained by a higher degree of specialization in households where fathers earn more. Therefore father in those households might not take all the weeks of parental leave that are reserved for them.

The next step would be to look at how earnings are related to number of weeks of each parent. A first approach would be to use a naive specification estimated for mothers and fathers separately ($j = m$ for mothers and $j = f$ for fathers):

$$p_{ji} = \delta_j w_{ji} + \beta_j X_{ji} + \epsilon_{ji}$$

In the above equation, p_{ji} is the amount of parental leave individual i takes, w_{ji} is that individual's pre-child wage, and X_{ji} is a set of controls. The issue with this naive approach is that households make parental leave decisions together. In particular, an increase in the amount of weeks one parent takes decreases the available weeks for the other parent. Therefore, the number of weeks one parent takes is a function of the number of weeks the other parent takes. Intuitively they should be substitutes. The true model should take this form (assuming a linear form for simplicity):

$$p_{mi} = \delta_m w_{mi} + \beta_m X_{mi} + \gamma_m p_{fi} + \epsilon_{mi}$$

$$p_{fi} = \delta_f w_{fi} + \beta_f X_{fi} + \gamma_f p_{mi} + \epsilon_{fi}$$

It is well known that such a simultaneous model is not identifiable as such. Without an instrument it is not possible to proceed further in the general case. However, focusing on fathers (which we are particularly interested in) there are two special cases where we can safely ignore substitution patterns. One case is when the mother is not in the labor force and therefore not eligible for parental leave benefits. In those cases, the father can freely choose the amount of weeks without worrying about restricting the mother's decision. The second case is when the mother takes the maximal amount of weeks available to her. In those cases, the choice of the father is reduced to choosing how much of the father's quota he will take.

Table 3.5 presents the results in those two special cases. I will focus first on the case where the mother takes the maximum available to her. In that case, higher income for the father will decrease the number of weeks he takes. However, the magnitude of the effect is very small with a 100 C$ per week increase in income only decreasing by 0.035 the expected number of days of parental leave. This small effect might be explained by the fact that there is not much variation in the data with most fathers

Table 3.4: The Effect of Income on Total Parental Leave

	Total weeks
Weekly Insurable Income Father	-0.0000442*
	(0.0000184)
Weekly Insurable Income Mother	0.000453***
	(0.0000176)
Self-employed Father	0.198***
	(0.0312)
Self-employed Mother	-1.360***
	(0.0364)
Age Father	0.00624***
	(0.00118)
Age Mother	0.0137***
	(0.00141)
Constant	53.11***
	(0.0369)
Observations	400349
R^2	0.009

Robust standard errors in parentheses, also contains year fixed effects
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

choosing 5 weeks of parental leave. The fact that the effect is negative indicates that for fathers the higher cost of parental leave in terms of foregone income outweighs the desire to consume more parental leave at higher income levels. This income effect can be observed for the first subgroup as the coefficient on the insurable income of mothers. An increase in that variable increases the total household income but keeps the cost of parental leave for the father constant. The effect when the mother is not in the labor force is much more substantial. A 100 C$ per week increase in income will decrease the expected number of days of parental leave by about 3.5 days. Here as well, the increased cost outweighs the income effect for the father's decision.

Table 3.5: The Effect of Income on Father's Parental Leave

	Mother takes max	Mother not in LF
Weekly Insurable Income Father	-0.0000683***	-0.00685***
	(0.00000414)	(0.000125)
Weekly Insurable Income Mother	0.0000271***	
	(0.00000396)	
Self-employed Father	0.00935	-1.017***
	(0.00671)	(0.231)
Age Father	-0.00267***	0.0917***
	(0.000214)	(0.00598)
Constant	4.891***	23.88***
	(0.00778)	(0.236)
Observations	296527	74528
R^2	0.004	0.040

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Conclusion

This paper shows that the introduction in 2006 of QPIP had a major impact on household parental leave behavior. While the reform impacted both fathers and mothers, the magnitude of the impact was much larger for fathers than for mothers. This paper also shows that the majority of households split the weeks of parental leave in a very specific way. This observed behavior suggests that the constraint introduced by the father's quota binds and that most households would prefer to specialize even further. Furthermore, richer households take more parental leave even though they face a higher cost (in terms of foregone income). However, when focusing only on fathers the opposite holds, as higher income fathers take less parental leave time.

In many ways this paper is still very preliminary. As discussed in the introduction much of the interest for this policy concerns the labor market outcomes. Therefore, using tax agency administrative data to look at the labor market effects is a crucial next step. Although I have not been able to access the tax data yet, what I know about its structure suggests a specific empirical strategy. The dataset only contains households that are residents of Quebec. However, it does contain new parents before and after the reform. As a result, to estimate the effect of the reform, the control group would be parents that had children just before the cutoff date. The large sample size and the arbitrary cutoff date lends itself to using a regression discontinuity design. Once again, the goal would be to see the effect in terms of labor market outcomes.

Another area for further work on this subject is building a model of parental leave decision-making. The goal of such a model would be to formalize some of the intuition discussed in this paper, especially in terms of the relation between income and parental leave weeks. A model would also make it possible to discuss the relative merits of counterfactual policies. The variation in prices introduced by the discontinuities in the replacement rate seem to be promising sources of variation to estimate preferences of households over parental leave.

REFERENCES

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, *105*(490), 493–505.

Ambler, K. (2015). Don't tell on me: Experimental evidence of asymmetric information in transnational households. *Journal of Development Economics*, *113*(100), 52–69.

Apps, P. F., & Rees, R. (1988). Taxation and the household. *Journal of Public Economics*, *35*(3), 355–369.

Ashraf, N. (2009). Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines. *American Economic Review*, *99*(4), 1245–1277.

Attanasio, O., & Lechene, V. (2014). Efficient Responses to Targeted Cash Transfers. *Journal of Political Economy*, *122*(1), 178–222.

Attanasio, O., & Lechene, V. (2002). Tests of the income pooling in household decisions. *Review of economic dynamics*, *5*(4), 720–748.

Attanasio, O. P., & Browning, M. (1995). Consumption over the life cycle and over the business cycle. *The American Economic Review*, *85*(5), pp. 1118–1137.

Attanasio, O. P., & Weber, G. (1995). Is consumption growth consistent with intertemporal optimization? evidence from the consumer expenditure survey. *Journal of Political Economy*, *103*(6), pp. 1121–1157.

Bargain, O., Lacroix, G., & Tiberti, L. (2018). Validating the Collective Model of Household Consumption Using Direct Evidence on Sharing, 46.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., & Pouliquen, V. (2015). Turning a shove into a nudge? A "labeled cash transfer" for education. *American Economic Journal: Economic Policy*, *7*(3), 86–125.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865.

Blundell, R., Chiappori, P.-A., Magnac, T., & Meghir, C. (2007). Collective labour supply: Heterogeneity and non-participation. *Review of Economic Studies*, *74*, 417–445.

Blundell, R., Chiappori, P.-A., & Meghir, C. (2005). Collective Labor Supply with Children. *Journal of Political Economy*, *113*(6), 1277–1306.

Bobonis, G. J. (2009). Is the allocation of resources within the household efficient? New evidence from a randomized experiment. *Journal of political Economy*, *117*(3), 453–503.

Boltz, M., Marazyan, K., & Villar, P. (2019). Income hiding and informal redistribution: A lab-in-the-field experiment in Senegal. *Journal of Development Economics*, *137*, 78–92.

Bronson, M. A. (2014). Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. *Unpublished manuscript*, *2*.

Bronson, M. A., & Mazzocco, M. (2018). Cohort size and the marriage market: Explaining nearly a century of changes in US marriage rates. *UCLA CCPR Population Working Papers*.

Brown, C., Calvi, R., & Penglase, J. (2018). Sharing the Pie: Undernutrition, Intra-Household Allocation, and Poverty. *SSRN Electronic Journal*.

Browning, M., & Chiappori, P.-A. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, *66*, 1241–1278.

Browning, M., Chiappori, P.-A., & Lewbel, A. (2013). Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power. *The Review of Economic Studies*, *80*(4), 1267–1303.

Browning, M., Bourguignon, F., Chiappori, P.-A., & Lechene, V. (1994). Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy*, *102*, 1067–96.

Browning, M., Hansen, L. P., & Heckman, J. J. (1999). Micro data and general equilibrium models. *Handbook of macroeconomics*, *1*, 543–633.

Browning, M., & Lusardi, A. (1996). Household saving: Micro theories and micro facts. *Journal of Economic literature*, *34*(4), 1797–1855.

Browning, M., & Meghir, C. (1991). The effects of male and female labor supply on commodity demands. *Econometrica*, *59*(4), pp. 925–951.

Cardenas, J. C., & Carpenter, J. (2008). Behavioural development economics: Lessons from field labs in the developing world. *The Journal of Development Studies*, *44*(3), 311–338.

Casanova, M. (2010). Happy together: A structural model of couples' joint retirement choices. *Unpublished Manuscript, Department of Economics, UCLA*.

Castilla, C., & Walker, T. (2013). Is Ignorance Bliss? The Effect of Asymmetric Information between Spouses on Intra-Household Allocations. *American Economic Review*, *103*(3), 263–268.

Cherchye, L., De Rock, B., Lewbel, A., & Vermeulen, F. (2015). Sharing rule identification for general collective consumption models. *Econometrica*, *83*(5), 2001–2041.

Cherchye, L., De Rock, B., & Vermeulen, F. (2007). The collective model of household consumption: A nonparametric characterization. *Econometrica*, *75*(2), 553–574.

Cherchye, L., De Rock, B., & Vermeulen, F. (2012). Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information. *American Economic Review*, *102*(7), 3377–3405.

Chiappori, P.-A., & Ekeland, I. (2009). The Microeconomics of Efficient Group Behavior: Identification. *Econometrica*, *77*(3), 763–799.

Chiappori, P.-A. (1988). Rational Household Labor Supply. *Econometrica*, *56*(1), 63–90.

Chiappori, P.-A. (1992). Collective Labor Supply and Welfare. *Journal of Political Economy*, *100*(3), 437–467.

Chiappori, P.-A., Fortin, B., & Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of political Economy*, *110*(1), 37–72.

Chiappori, P.-A., & Mazzocco, M. (2017). Static and intertemporal household decisions. *Journal of Economic Literature*, *55*(3), 985–1045.

Chiappori, P.-A., & Meghir, C. (2015). Chapter 16 - Intrahousehold Inequality. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution* (pp. 1369–1418). Elsevier.

Cho, Y., & Ruthbah, U. (2018). *Does Workfare Work Well? The Case of the Employment Generation Program for the Poorest in Bangladesh*. The World Bank.

Clavet, N.-J., Corneau-Tremblay, N., & Lacroix, G. (2016). *Evaluation des retombées du régime québécois d'assurance parentale*. Québec, Département d'économique, Université Laval.

Cools, S., Fiva, J. H., & Kirkebøen, L. J. (2015). Causal effects of paternity leave on children and parents. *The Scandinavian Journal of Economics*, *117*(3), 801–828.

Dagnelie, O., & Lemay-Boucher, P. (2012). Rosca Participation in Benin: A Commitment Issue*: Rosca participation in Benin. *Oxford Bulletin of Economics and Statistics*, *74*(2), 235–252.

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, *104*(7), 2049–74.

Dahl, G. B., Løken, K. V., Mogstad, M., & Salvanes, K. V. (2016). What is the case for paid maternity leave? *Review of Economics and Statistics*, *98*(4), 655–670.

Dercon, S., & Krishnan, P. (2000). In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia. *Journal of Political Economy*, *108*(4), 688–727.

Doepke, M., & Tertilt, M. (2019). Does female empowerment promote economic development? *Journal of Economic Growth*, *24*(4), 309–343.

Donni, O. (2009). A simple approach to investigate intrahousehold allocation of private and public goods. *The Review of Economics and Statistics*, *91*(3), 617–628.

Dubois, P., & Ligon, E. A. (2011). Incentives and nutrition for rotten kids: Intrahousehold food allocation in the Philippines.

Duflo, E., & Udry, C. (2004). Intrahousehold resource allocation in Cote d'Ivoire: Social norms, separate accounts and consumption choices. *National Bureau of Economic Research*.

Dunbar, G. R., Lewbel, A., & Pendakur, K. (2013). Children's resources in collective households: Identification, estimation, and an application to child poverty in Malawi. *American Economic Review*, *103*(1), 438–71.

Dunbar, G. R., Lewbel, A., & Pendakur, K. (2019). Identification of random resource shares in collective households without preference similarity restrictions. *Journal of Business & Economic Statistics*, 1–20.

Ekberg, J., Eriksson, R., & Friebel, G. (2013). Parental leave—a policy evaluation of the swedish "daddy-month" reform. *Journal of Public Economics*, *97*, 131–143.

Fong, Y., & Zhang, J. (2001). The identifiability of unobservable independent and spousal leisure. *Journal of Political Economy*, *109*(1), 191–202.

Gayle, G.-L., & Shephard, A. (2019). Optimal taxation, marriage, home production, and family labor supply. *Econometrica*, *87*(1), 291–326.

Gemici, A., & Laufer, S. (2012). Marriage and cohabitation. *Working Paper, NYU*.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Green, J. R., & Laffont, J.-J. (1986). Partially Verifiable Information and Mechanism Design. *The Review of Economic Studies*, *53*(3), 447.

Greene, W. H. (2003). *Econometric analysis* (5th ed). Prentice Hall
page 370.

Greenwood, J., Guner, N., & Knowles, J. A. (2003). More on marriage, fertility, and the distribution of income. *International Economic Review*, *44*(3), 827–862.

Gross, D. B., & Souleles, N. S. (2002). Do liquidity constraints and interest rates matter for consumer behavior? evidence from credit card data. *The Quarterly journal of economics*, *117*(1), 149–185.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American economic review*, 622–641.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, *86*(6), pp. 971–987.

Hall, R. E., & Mishkin, F. S. (1982). The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. *Econometrica*, *50*(2), 461–481.

Haushofer, J., & Shapiro, J. (2013). Household response to income changes: Evidence from an unconditional cash transfer program in Kenya. *Massachusetts Institute of Technology*, *24*(5), 1–57.

Hoel, J. B. (2015). Heterogeneous Households A within-subject test of asymmetric information between spouses in Kenya. *Journal of economic behavior & organization*, *118*, 123–135.

Khandker, S. R., & Mahmud, W. (2012). *Seasonal Hunger and Public Policies: Evidence from Northwest Bangladesh*. The World Bank.

Klein, M. J., & Barham, B. L. (2018). Point Estimates of Household Bargaining Power Using Outside Options.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Lecocq, S., & Robin, J.-M. (2015). Estimating Almost-ideal Demand Systems with Endogenous Regressors. *The Stata Journal*, *15*(2), 554–573.

Ligon, E. A. (2011). Dynamic Bargaining in Households (with an Application to Bangladesh). *SSRN Electronic Journal*.

Low, H., Meghir, C., Pistaferri, L., & Voena, A. (2018). Marriage, labor supply and the dynamics of the social safety net. *National Bureau of Economic Research*.

Malapit, H. J. L. (2012). Why do spouses hide income? *The Journal of Socio-Economics*, *41*(5), 584–593.

Manser, M., & Brown, M. (1980). Marriage household decision-making: A bargaining analysis. *International Economic Review*, *21*, 31–44.

Mazzocco, M. (2004). Saving, risk sharing, and preferences for risk. *American Economic Review*, *94*(4), 1169–1182.

Mazzocco, M. (2007). Household intertemporal behaviour: A collective characterization and a test of commitment. *The Review of Economic Studies*, *74*(3), 857–895.

McElroy, M. B., & Horney, M. J. (1981). Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, *22*, 333–349.

Meghir, C., Mommaerts, C., Mobarak, A. M., & Morten, M. (n.d.). Migration and Informal Insurance, 78.

Meghir, C., & Weber, G. (1996). Intertemporal nonseparability or borrowing restrictions? A disaggregate analysis using a U.S. consumption panel. *Econometrica*, *64*(5), pp. 1151–1181.

Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, *37*(4), 1009–1059.

Pollak, R., & Wales, T. J. (1981). Demographic Variables in Demand Analysis. *Econometrica*, *49*(6), 1533–51.

Prescott, E. C. (1986). Theory ahead of business-cycle measurement. *Carnegie-Rochester conference series on public policy*, *25*, 11–44.

Robinson, J. (2012). Limited Insurance within the Household: Evidence from a Field Experiment in Kenya. *American Economic Journal: Applied Economics*, *4*(4), 140–164.

Sargent, T. J. (1978). Rational expectations, econometric exogeneity, and consumption. *Journal of Political Economy*, *86*(4), pp. 673–700.

Schultz, P. (1990). Testing the neoclassical model of family labor supply and fertility. *Journal of Human Resources*, *25*(4), 599–634.

Squires, M. (n.d.). Kinship Taxation as an Impediment to Growth: Experimental Evidence from Kenyan Microenterprises, 74.

Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of Human Resources*, *25*, 635–664.

Voena, A. (2015). Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples? *American Economic Review*, *105*(8), 2295–2332.

Waldfogel, J. (1997). The effect of children on women's wages. *American sociological review*, 209–217.

Woods, J. E. (1979). A note on Hicks' Composite Commodity Theorem. *Zeitschrift für Nationalökonomie*, *39*(1-2), 185–188.

Zeldes, S. P. (1989). Consumption and liquidity constraints: An empirical investigation. *Journal of Political Economy*, *97*(2), pp. 305–346.