UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2017/05

Cambridge Working Papers in Economics: 1711

AGGREGATING ELASTICITIES: INTENSIVE AND EXTENSIVE MARGINS OF FEMALE LABOUR SUPPLY

Orazio Attanasio   Peter Levell   Hamish Low   Virginia Sanchez-Marcos

We estimate labour supply elasticities at the micro level and show what we can learn from possibly very heterogeneous elasticities for aggregate behaviour. We consider both intertemporal and intratemporal choices, and identify intensive and extensive responses in a consistent life cycle framework, using US CEX data. There is substantial heterogeneity in how individuals respond to wage changes at all margins, both due to observables, such as age, demographics, and wealth, as well as to unobservable tastes for leisure. We estimate the distribution of Marshallian elasticities for hours worked to have a median value of 0.18, and corresponding Hicksian elasticities of 0.54 and Frisch elasticities of 0.87. At the 90th percentile, these values are 0.79, 1.16, and 1.92. Responses at the extensive margin are important, explaining about 54% of the total labour supply response for women under 30, although this importance declines with age. We also show that aggregate elasticities are cyclical, being larger in recessions and particularly so in long recessions. This heterogeneity at the micro level means that the aggregate labour supply elasticity is not a structural parameter: any macro elasticity will depend on the demographic structure of the economy as well as the distribution of wealth and the particular point in the business cycle.

Aggregating Elasticities: Intensive and Extensive Margins of Female Labour Supply *

Orazio Attanasio[†] Peter Levell [‡] Hamish Low [§] Virginia Sánchez-Marcos[¶]

January 31, 2017

Abstract

We estimate labour supply elasticities at the micro level and show what we can learn from possibly very heterogeneous elasticities for aggregate behaviour. We consider both intertemporal and intratemporal choices, and identify intensive and extensive responses in a consistent life cycle framework, using US CEX data. There is substantial heterogeneity in how individuals respond to wage changes at all margins, both due to observables, such as age, demographics, and wealth, as well as to unobservable tastes for leisure. We estimate the distribution of Marshallian elasticities for hours worked to have a median value of 0.18, and corresponding Hicksian elasticities of 0.54 and Frisch elasticities of 0.87. At the 90th percentile, these values are 0.79, 1.16, and 1.92. Responses at the extensive margin are important, explaining about 54% of the total labour supply response for women under 30, although this importance declines with age. We also show that aggregate elasticities are cyclical, being larger in recessions and particularly so in long recessions. This heterogeneity at the micro level means that the aggregate labour supply elasticity is not a *structural* parameter: any macro elasticity will depend on the demographic structure of the economy as well as the distribution of wealth and the particular point in the business cycle.

JEL Codes: J22, D91

Keywords: labour supply elasticities, heterogeneity, aggregation, non-separability

*We are grateful for a number of useful conversations with Richard Blundell, Guy Laroque, Costas Meghir, Richard Rogerson and Tom Sargent. We received several useful comments from different seminar audiences and during presentation at the NBER Summer Institute and the Society for Economic Dynamic Conference. Attanasio's research was partially funded by an ESRC Professorial Fellowship and by the ESRC Centre for the Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies. Sánchez-Marcos thanks Spanish MCYT for Grant ECO2009-09614.

[†]UCL, IFS and NBER

[‡]IFS and UCL

[§]University of Cambridge and IFS

[¶]Universidad de Cantabria

1 Introduction

The size of the elasticity of labour supply to changes in wages has been studied for a long time. Recent debates have focused on the perceived discrepancy between estimates coming from micro studies, which, with a few exceptions, point to relatively small values of such an elasticity, and the assumptions made in macro models, which seem to need relatively large values. Keane and Rogerson (2015) and Keane and Rogerson (2012) survey some of these issues and the papers by Blundell et al. (2011), Ljungqvist and Sargent (2011) and Rogerson and Wallenius (2009) contain some alternative views on the debate. To reconcile the micro evidence and the assumptions made in macroeconomics, much attention has been given to the distinction between the extensive and intensive margins of labour supply, see, in particular, Chetty (2012) and Chetty et al. (2011). Perhaps surprisingly, in this debate, aggregation issues and the pervasive and complex heterogeneity that characterise labour supply behaviour have not been given much attention. This paper aims to fill this gap, while making some original methodological contributions and presenting new empirical evidence.

Preferences for consumption and leisure are bound to be affected in fundamental ways by family composition, health status, fertility, as well as unobserved tastes 'shocks', and so heterogeneity in labour supply elasticities in these dimensions is something to be expected. The key issue, however, is how significant this heterogeneity is and whether it is important at the aggregate level: does it make any sense to talk about *the* elasticity of labour supply as a *structural* parameter? Aggregation issues are likely to be relevant both for the intensive and extensive margin, as we show. Labour supply elasticities vary much in the cross section and, importantly, over the business cycle.

In this paper, we address these issues focusing on female labour supply. Our approach consists in taking a relatively standard life cycle model of labour supply to the data. Whilst the essence of the model is relatively simple, we stress two elements that are important for our analysis and that make our contribution novel. First, we consider all the relevant intertemporal and intratemporal margins and choices simultaneously; in particular, consumption and saving as well as participation and hours of work. Second, we specify a flexible utility function that allows for substantial heterogeneity, fits the data well and, at the same time allows us to make precise quantitative statements. These elements are important because they allow us to address directly the interaction between extensive and intensive margins and to evaluate empirically the importance of aggregation issues and to calculate both micro and macro elasticities.

In evaluating aggregate labour supply elasticities, it is necessary to specify the whole economic environment because, as noted by Chang and Kim (2006), the aggregate response depends on the distribution of reservation wages. On the other hand, our key methodological contribution is to stress that a number of key components of the model can be estimated using considerably weaker assumptions while maintaining consistency with the overall model structure. We separate our estimation into three steps and specify explicitly what assumptions are needed at each step. The first step identifies the within-period preferences over consumption and labour supply at the intensive margin; the second step identifies intertemporal preferences; the third step identifies the fixed costs and full economic environment that drive the participation decisions over the life cycle. In the first step, Marshallian and Hicksian elasticities at the intensive margin can be computed from the parameters of the Marginal Rate of Substitution between consumption and leisure, which can be estimated using only static conditions and without taking a stance on intertemporal allocations or participation decisions. This means that the estimates of these elasticities we get are robust to further specific assumptions made on the economic environment.[1] Analogously, to compute the Frisch elasticity, we rely on estimates of the elasticity of intertemporal substitution that can be derived from the Euler equation for consumption, again without taking a stance on participation. Finally, to assess and characterise behaviour at the extensive margin, we need to specify the model fully. In this step, we characterise the behaviour of such a model by calibrating its key parameters to a number of life cycle moments. Our approach effectively uses two stage budgeting ideas and conditional demands.

While throughout the paper we make specific assumptions about the shape of the utility function, we use flexible specifications to allow for observed and unobserved heterogeneity in tastes. In particular, we allow many observed variables to affect the intratemporal and intertemporal margins while at the same time allowing for possible non-separability of consumption and leisure. Our specification of preferences is much more flexible than the ones that are in general considered in the literature and this is important. Classic papers in the micro literature (such as Heckman and Macurdy (1980)) imply a strong relationship between the Frisch intertemporal elasticity and the intratemporal Marginal Rate of Substitution conditions. This forces a strict relationship between intraperiod and intertemporal conditions which our approach avoids. In the macro literature, most papers impose additive separability between consumption and leisure, and isoleastic, homothetic preferences that conform to the restrictions for balanced growth, as in Erosa et al. (2016). This assumption is predicated on the perceived need to work with models that match historical trends showing steady secular increases in real wages with little change in aggregate hours. Browning et al. (1999) already noted that the fact that the historical trend for aggregate hours is roughly constant hides a large decrease for males and an increase for females. Here, we show that the isoelastic specification for consumption and hours is strongly rejected by the data. The challenge, therefore, is to work with specifications that admit much more heterogeneity and changes over time.

When bringing the model to the data, we are explicit about what variation in the data identifies each component of the model. When considering the within-period MRS condition, we do not use

[1] The MRS condition hold as an equality only for women who work. This implies that, empirically, we need to address the issue of participation. However, this can be done with a reduced form that is consistent with the model we are studying as well as other more general models.

the variation in individual wages, which is likely to be related to individual characteristics correlated with unobserved taste heterogeneity. Instead, we use aggregate variability that is driven by aggregate shocks such as policy reforms. This makes our approach in this step similar to Blundell et al. (1998). In our second step, estimating the Euler equation for consumption, we take into account the presence of unobserved taste shocks and the fact that the lack of longitudinal data forces us to work with synthetic cohort data; our approach here is similar to Attanasio and Weber (1995). Finally, in our third step, we estimate the full life-cylce model and explicitly aggregate individual behaviour.

Estimates of the size of the elasticity of labour supply in the literature vary considerably even for women. Different authors have used different identification strategies, different specifications and different data sources. Our estimates, at the median, are not too different from some of the estimates in the literature. In particular, on the intensive margin, we obtain a Marshallian median elasticity of 0.18, with the corresponding Hicksian elasticity considerably larger at 0.54, indicating a sizeable income effect. For the same median household, the Frisch elasticity for hours is 0.87. At the same time, we document considerable variation in the size of the estimated elasticities in the cross section. The Marshallian, for instance, varies between -0.14 at the 10th percentile and 0.79 at the 90th percentile.

Using the entire model, we can aggregate explicitly individual behaviour and study aggregate elasticities that correspond to the concept used in the macro literature. We find an important role for the extensive margin in generating aggregate movements in labour supply. And, even at the extensive margin, we find a considerable amount of heterogeneity in the cross section, driven by age, the number of children and wealth. Most importantly in linking the micro and macro analysis of labour supply, we show that what we call the 'macro' elasticity changes considerably over the business cycle, and is typically larger in recessions. Moreover, it gets larger in longer recessions. To the best of our knowledge, changes in the elasticity over the business cycles have never been discussed.

The closest macroeconomic paper to ours is Erosa et al. (2016), who have similar aims of building aggregate elasticities from individual behaviour over the life cycle, and of distinguishing the intensive and extensive margins using a fully specified life cycle model. The focus of their paper is on male labour supply, rather than the female labour supply responses. A second related paper is Guner et al. (2012), who model heterogeneous married and single households with a female extensive margin and a male and female intensive margin. Their focus is on evaluating different reforms of the US tax system and they abstract from wage uncertainty. Both papers operate with very specific preference specifications. We discuss the extent to which our results differ from these papers in the conclusions.

We use the fully specified model to run two experiments: in the first, we evaluate the labour supply response to an anticipated temporary changes in wages; in the second, we evaluate the response to an unanticipated change in the entire life cycle wage profile. The first experiment captures the impact of a temporary tax cut: wages are low today, but known to increase in the next period, and the marginal utility of wealth is likely to be unaffected. The second captures the impact of a permanent tax cut which will change the marginal utility of wealth. Without an extensive margin, the response to the first experiment would be captured by the Frisch and the response to the second can be approximated by the Marshallian. Introducing the extensive margin doubles the size of the response in the first experiment, and is particularly important at younger ages when non-participation because of children is prevalent. Further, the extensive margin is especially important when we simulate responses to tax changes in recessions: responses are higher in recessions, and the response increases as the recession persists. On the other hand, the extensive margin response is much more muted in the second experiment of a shift in the entire life cycle wage profile.

In comparing our estimates to those available in the literature, we investigated extensively what drives, in our data, differences in results. As it turns out, the size of the elasticities we obtain does not depend much on the set of instruments we use, nor on the fact that we allow education levels to affect preferences. Instead, and more prosaically, the size of the estimates depend on the specific estimator and normalisation used. When using standard IV or GMM methods, and when the coefficient on wages in the equation that relates wages to the MRS is normalized to one, we typically obtain very large estimates. Instead, when the coefficient on consumption or hours worked is normalised to one, we obtain much smaller estimates. In our baseline estimation, we use methods whose results are not affected by the normalization used. In particular, we use a method proposed by Fuller, which is a generalization of a LIML approach.

Our exercise is not without important caveats. In much of our analysis, we do not consider the effect of tenure and experience on wages. Such effects can obviously be important, as labour supply choices may change future wages and, therefore, future labour supply behaviour, as stressed by Imai and Keane (2004). In section 7, we extend our analysis to introduce returns to experience on the extensive margin. Introducing returns only on the extensive margin means within-period allocations at the intensive margin are unaffected. By contrast, if the return to experience operates on the number of hours (rather than only on participation), we would need to change our analysis substantially.

The rest of the paper is organized as follows. In section 2, we outline the assumptions of the life cycle framework. We split the sets of assumptions into three groups, corresponding to the three steps of the estimation process to make explicit what the minimum set of assumptions are needed at each stage. We also show how the preference parameters can be mapped into static and intertemporal elasticities, and discuss the meaning of the different elasticities. In section 3 we explain the three steps of our empirical strategy to identify the preference parameters, that is, using intraperiod first order conditions, intertemporal first order conditions and full structural estimation. Section 4 describes the data and provides some descriptive statistics. Section 5 presents and discusses the parameter estimates. In section 6 we report the implications of our estimates for labour supply elasticities, distinguishing between Marshallian, Hicksian and Frisch elasticities. We also report responses on the extensive margin, aggregate responses and, more generally, the aggregation issues that are central to our argument. Section 7 extends the analysis to include returns to experience and section 8 concludes.

2 A life cycle model of female labour supply

To study both the intensive and the extensive margin elasticity of female labour supply, we use a rich model of female labour supply choices embedded in a life cycle framework. A unitary household makes choices about consumption and female labour supply, given exogenous processes for male earnings and female wages and an intertemporal budget constraint. Both the intensive and extensive margins are meaningful because of the presence of fixed costs of going to work related to family composition and/or because of the presence of preference costs specifically related to participation.

2.1 Model structure and assumptions

Within the life cycle model we use, it is useful to characterise individual choices in three steps. Using two-stage-budgeting, we first model the within-period allocation between consumption and hours of work. We then characterise the intertemporal allocation of consumption and hours of work; and, finally, the dynamic extensive margin and the evolution of labour force participation. These three steps map into three stages of estimation, as outlined in section 3 below, with increasing structure being required on the underlying model at each stage. Corresponding to each step there is a set of additional assumptions characterising the underlying model of behaviour. We specify and discuss the assumptions for each step in turn.

2.1.1 Intratermporal choice

A1: Within Period Utility. We consider married couples, who maximise the utility of the household and choose consumption and female labour supply within each period. In making the within-period decision, households take total spending within the period to be given. The allocation of total spending between periods is considered in the next stage of the model. We define the within-period utility for each individual household h across consumption and hours of leisure per week:[2]

$$M_{h,t}\left(c_{h,t}, l_{h,t}; z_{h,t}, \chi_{h,t}\right) = \left(\frac{(c_{h,t}^{1-\phi} - 1)}{1-\phi} + \left(\alpha_{h,t}(z_{h,t}, \chi_{h,t})\right) \frac{(l_{h,t}^{1-\theta} - 1)}{1-\theta} \right) \tag{1}$$

where the subscripts h and t index the individual household and age respectively, $z_{h,t}$ is a vector of demographic variables (which include education, age and family composition) that are observable in

[2]We adopt a parametric rather than a more flexible functional form for utility. As we will see this leads to a log-linear MRS condition. Relative to a non-parametric approach, this imposes more restrictions on behaviour. However it guarantees integrability of the MRS conditions and allows us to readily verify that quasi-concavity of the utility function is satisfied. Quasi-concavity, requires that the MRS between consumption and leisure is decreasing in leisure and increasing in consumption. After estimating the relevant parameters, these conditions can be checked empirically.

the data; and the term $\chi_{h,t}$ represents a taste shifter that is unobservable in the data, but is cohort specific and is known to the individual. The function $\alpha_{h,t}$ is specified as:

$$\alpha_{h,t} = \exp(\psi_0 + \psi_z z_{h,t} + \chi_{h,t}) \tag{2}$$

where ψ_0 and ψ_z are parameters, as are ϕ and θ.

In thinking about within-period decisions, there is no difference to the within-period allocation from using the specification (1) or any monotonic transformation of M such as $f\left(M_{h,t}\left(c_{h,t}, l_{h,t}\right)\right)$. In other words, introducing non-separability between consumption and leisure through such a transformation f has no implications for within-period allocations. An implication of this consideration is that non separability between consumption and leisure cannot be identified from within-period choices.

This specification is more general than a standard Cobb-Douglas or a standard CES. A Cobb-Douglas specification would require $\phi = \theta = 1$ and imposes a unitary Marshallian elasticity of consumption with respect to the wage and zero elasticity for hours of work. This restriction would limit substantially the amount of heterogeneity in elasticities. For balanced growth (in female labour supply) we would require $\phi = 1$. A standard CES would require $\phi = \theta$ and constrains the elasticity of substitution for leisure and consumption. As we discuss in section 3 below, the specification in equation (1) is the most general specification for within-period preferences that gives rise to a linear in parameters estimating equation.

A2: Within Period Budget Constraint. Total within-period spending is allocated to consumption and to leisure.

$$x_{h,t} = x^0_{h,t} + w^f_{h,t} H = c_{h,t} + w^f_{h,t} l_{h,t} \tag{3}$$

where $x^0_{h,t}$ is total unearned income, including husband's earnings and the difference in wealth between the start and the end of the period and subtracts off any fixed cost of work.

A3: Period Length. One period in the model is one quarter. Households choose typical hours of work each week and then this is constant across weeks within the quarter, to give within-period hours of work. The *extensive margin* is the decision whether or not to work that quarter. The *intensive* margin is how many hours to work in a typical week. This assumption means we are not allowing individuals to choose how many weeks to work in a quarter. This restriction is driven by data limitations since we do not observe the number of weeks per quarter that an individual works, we only observe typical hours per week and number of weeks per year. However, we provide empirical support for our approach in section 4.3.

A4: Returns to Experience. Wages for women do not depend on the history of labour supply and evolve exogenously. This assumption means that decisions on labour supply today do not have a direct effect on the continuation value for the future. In other words, the only linkage across periods is through the decision about total within-period spending. This assumption is necessary in order to invoke two stage budgeting and focus on the within-period problem without considering the intertemporal allocation. We relax the assumption that there are no returns to experience in section 7 where we distinguish the cases where returns to experience depend on participation and where returns depend on hours worked. Our estimation approach will go through if returns to experience operate through the participation margin rather than through the hours of work margin.

2.1.2 Intertemporal choice at the intensive margin

A5: Intertemporal Utility. Households maximise lifetime expected utility of the household:

$$\max_{c,l} E_t \sum_{j=0}^{T} \beta^j u\left(c_{h,t+j}, l_{h,t+j}, P_{h,t+j}; z_{h,t+j}, \chi_{h,t+j}, \zeta_{h,t+j}\right) \tag{4}$$

$$u\left(c_{h,t}, l_{h,t}\right) = \frac{M_{h,t}^{1-\gamma}}{1-\gamma} \exp(\pi z_{h,t} + \xi P_{h,t} + \zeta_{h,t}) \tag{5}$$

The variable $P_{h,t}$ is an indicator of the woman's labour force participation, which can affect utility over and above the effect of hours worked. As with the instantaneous utility function in (1), $z_{h,t}$ is a vector of observable characteristics that are observable in the data. The term $\zeta_{h,t}$ represents taste shifters that affect intertemporal choices and are unobservable in the data. There are cohort specific and known to the individual: this taste shifter is different from but not necessarily uncorrelated with $\chi_{h,t}$ which affects intratemporal choices. Notice that the observable characteristics $z_{h,t}$ that appear in equations (2) and (5) need not be the same.

Our specification allows for non-separability between consumption and leisure both at the intensive and extensive margin, and for the effect of demographic variables and unobserved taste shocks to affect utility. Further, it does not impose that the elasticities of substitution for leisure and for consumption are constant.

Notice that if the estimate of γ is equal to 0, then the utility function collapses to the additively separable form and the elasticity of intertemporal substitution of consumption would equal ϕ and could be estimated from the within-period MRS condition alone.

The general specification of utility we use allows substantial heterogeneity across individuals in intratemporal and intertemporal preferences, across the intensive and extensive margins. Heterogeneity arises partly because elasticities will differ by observable characteristics: by education and the presence of children; partly because elasticities differ at different levels of consumption and hours of work. Finally, we allow for unobserved heterogeneity in estimates of the intensive margin elasticities through χ as well as in the extensive margin through ζ. In other words, we allow the data to pin down where heterogeneity will be more important. If we had adopted a simple Cobb-Douglas specification, there is much less scope for heterogeneity at the intensive margin, and heterogeneity would have come through the extensive margin and the distribution of reservation wages.

A6: Intertemporal Budget Constraint. Maximisation is subject to an intertemporal budget constraint of the form:

$$A_{h,t+1} = (1 + r_{t+1}) \left(A_{h,t} + \left(w^f_{h,t} \left(H - l_{h,t} \right) - F \left(a_{h,t} \right) \right) P_{h,t} + w^m_{h,t} \bar{h} - c_{h,t} \right) \tag{6}$$

where $A_{h,t}$ is the beginning of period asset holding, r_t is the interest rate, F the fixed cost of work, dependent on the age of children $a_{h,t}$. Female wages are given by $w^f_{h,t}$, and husband's wages are given by $w^m_{h,t}$, with fixed husband hours of work, $\bar{h}$. Our assumption that husband hours of work are fixed may have implications for our estimated response of the female labor supply to aggregate shocks. In principle, it is reasonable to expect that male hours would go in the same direction as female hours and as a consequence there would be an income effect that would possibly mitigate the female responses.

A7: Borrowing Constraints. There are no explicit borrowing constraints but households cannot go bankrupt. This means in each period, households are able to borrow against the minimum income they can guarantee for the rest of their lives. Households have no insurance markets to smooth aggregate or idiosyncratic shocks and must rely on self-insurance.

2.1.3 Choice at the extensive margin

To model the extensive margin explicitly, we need to specify the full economic environment individuals face, including details of the cost of work and all relevant stochastic processes. The set of assumptions made here are not needed to bring the first two steps to the data, which relies on first-order conditions.

A8: The Fixed Cost of Work. We assume that the cost of work, F, depends on $a_{h,t}$, the age of the youngest child. We denote the child care units needed by $G(a_{h,t})$ and the price of each unit of child care by p. Therefore, the total fixed cost faced by a household when women participate in the labour market is given by

$$F(a_{h,t}) = pG(a_{h,t}) + \bar{F} \tag{7}$$

The fixed cost of work is deterministic and fully known. The presence of fixed costs of going to work and discrete utility costs introduces the possibility that some women will decide not to work at all, especially at low levels of productivity. By the same token, it will be unlikely that women who do choose to work will work only very few hours.

A9: Wage Processes. Men always work and male wages are given by:

$$\ln w^m_{h,t} \quad = \quad \ln w^m_{h,0} + \iota^m_1 t + \iota^m_2 t^2 + v^m_{h,t} \tag{8}$$

Female wages are instead given by the following process:

$$\ln w^f_{h,t} = \ln w^f_{h,0} + \ln e^f_{h,t} + v^f_{h,t} \tag{9}$$

where $e^f_{h,t}$ is the level of female human capital at the start of the period. In our baseline specification, human capital does not depend on the history of labour supply and is assumed to evolve exogenously according to[3]

$$\ln e^f_t = \iota^f_1 t + \iota^f_2 t^2 \tag{10}$$

We assume that wage rates do not depend on the number of hours worked, therefore ruling out part-time penalties. This assumption, for women, is consistent with what we observe in our data and with other US-based studies (Hirsch (2005); Aaronson and French (2004)).

There are initial distributions of wages for both men and women, $w^m_{h,0}$, $w^f_{h,0}$. Female and male wages are subject to permanent shocks, $v^f_{h,t}$ and $v^m_{h,t}$, that are positively correlated. We assume

$$\begin{aligned} v_{h,t} &= v_{h,t-1} + \xi_{h,t} && (11)\\ \xi_{h,t} &= (\xi^f_{h,t}, \xi^m_{h,t}) \sim N\left(\mu_\xi, \sigma^2_\xi\right) && (12)\\ \mu_\xi &= \left(-\frac{\sigma^2_{\xi^f}}{2}, -\frac{\sigma^2_{\xi^m}}{2}\right) \text{ and } \sigma^2_\xi = \begin{pmatrix} \sigma^2_{\xi^f} & \rho_{\xi^f,\xi^m} \\ \rho_{\xi^f,\xi^m} & \sigma^2_{\xi^m} \end{pmatrix} \end{aligned}$$

A10: Interest Rates. In the first two steps, we do not have to specify a process for the interest rate. In the second stage, we use variation in the interest rate to estimate intertemporal preference parameters. In neither of the first two stages, do we need to make any assumption on the assets available to individuals to move resources over time. Even when estimating intertemporal preferences, we need a household not to be at a corner for the specific asset we consider, regardless of the presence of frictions in other asset markets. When calibrating the model in the third step, we assume a constant interest rate to keep the model tractable.

A11: Fertility. Fertility is deterministic and children are fully anticipated, although women differ in their age at childbirth. In reality, there is clearly some degree of uncertainty in the realisation of households fertility decisions, however we have not made fertility a full-stochastic outcome as that would have increased the numerical complexity of the problem substantively.

A12: Marital Status. There is no uncertainty about marital status or divorce.

A13: Job Market Frictions. There is no job destruction or search. All periods of not-working are by choice in light of the offered wage.

Given the dynamic problem we have just described, individual households will make decisions taking the stochastic processes above as given. When considering aggregation, we need to take a stand on the degree of correlations in the shocks different households receive. We assume that households are

[3]In section 7, we generalise this process to allow for returns to experience.

subject to both idiosyncratic and aggregate shocks and so the shocks that affect individual households at a given point in time are correlated. However, from the household's perspective, they do not distinguish aggregate and idiosyncratic shocks and condition their future expectations only on their own observed wage realisations. Our framework is not a general equilibrium one: we do not construct the equilibrium level of wages (and interest rates). However, we study aggregate female labour supply and its elasticity to wages. We do so by simulating a large number of households and aggregating explicitly their behaviour.

2.2 Marshallian, Hicksian and Frisch elasticities

As argued by Blundell and MaCurdy (1999) and, more recently by Keane (2011), estimates of the 'wage elasticity' may refer to different quantities depending on the type of variation in wages one is considering. On the one hand, one can consider the effect of changes in the entire wage structure, as induced, for instance, by a permanent changes in labour income taxation (or in the comparison between different countries). On the other, one can consider short term variations in wages, such as those one observe over the business cycle, akin to what Blundell and MaCurdy (1999) and MaCurdy (1983) define 'evolutionary' wage changes; a distinction also made by Chetty (2012) and Chetty et al. (2013). These different type of variations in wages in the data can be mapped in different theoretical concepts. The size of changes in labour supply induced by evolutionary wage changes is related to the size of the Frisch (or marginal utility of wealth constant) elasticity, while the size of changes induced by permanent shifts to the wage structure are determined by the size of Hicksian or Marshallian elasticities, depending on whether the changes in wages are compensated or not.[4]

More specifically, in a dynamic model, a Marshallian elasticity describes how hours of work within a period change holding end-of-period savings constant, whereas a Hicksian response conditions on utility within the period. As suggested by Keane (2009), an alternative representation of the Hicks elasticity can be given considering a tax change with a lump-sum transfer, keeping life cycle wealth constant.[5] In such a situation, the Marshallian elasticity would describe the change in labour supply if the tax change is not compensated. Therefore, if one wants to think about the implications for labour supply of permanent changes in taxes, the Marshallian and Hicksian elasticities are the relevant concepts, but both miss the possibility that savings change in response to the change in taxes.

Following the change in the structure of wages (possibly induced by changes in taxes), resources may be reallocated over time through changes to the time path of hours of work changing or through

[4]Blundell and MaCurdy (1999) and Keane (2011) discuss clearly how the concepts of Marshallian and Hicksian elasticities, which are typically derived within the framework of a static model can be put within the framework of a dynamic life cycle model through the machinery of two-stage budgeting, as developed by Gorman (1959) and applied to labour supply by MaCurdy (1981), MaCurdy (1983) and Blundell and Walker (1986)

[5]This concept of a Hicks elasticity is used in Keane and Rogerson (2011) and in Chetty (2012). It is equivalent to the static concept under the assumption that resources are freely transferable between periods and preferences are separable between consumption and leisure. Alternatively, it is equivalent if preferences are quasilinear, in which case the Marshallian, Hicksian and Frisch elasticities coincide.

changes to the time path of the marginal utility of wealth, or both. The Frisch elasticity captures the change over time in hours worked in response to the anticipated evolution of wages, with the marginal utility of wealth unchanged because the wage change conveys no new information.[6] This is then the right concept if one wants to think about the implications of changes in wages over the business cycle. Similarly, it is the right concept to use if we are thinking about temporary changes to taxation.

In each of these cases, the labour supply response can be thought of in terms of the intensive (hours) or the extensive (participation) margin. At the individual level, an elasticity is easily defined when thinking of the intensive margin, while the same concept is somewhat vaguer when thinking of the extensive margin, especially when thinking of the Frisch elasticity that is supposed to keep the marginal utility of wealth constant. For a macroeconomist, the next step is to think of how these individual responses are reflected in changes in employment and hours of work. Indeed, in the case of the extensive margin, one can think of the impact that a change in wages has on the fraction of individuals that change their participation status, given the distribution of state variables. In this sense, the consideration of the extensive margin brings to the forefront aggregation issues that have not figured prominently in the discussion of labour supply elasticities. Aggregate participation responses to an aggregate shock are bound to depend on the distribution of state variables in the cross section. As we discuss below, aggregation issues can also be relevant for the intensive margin.

2.3 Marginal rate of substitution

We first consider the problem of allocating resources between consumption and female leisure within each period. The first block of assumptions in section 2.1.1 imply standard two-stage budgeting because of separability over time in the utility function and separability over time in the budget constraint. We do not need to make any assumptions about the intertemporal budget constraint, borrowing constraints, the nature of fixed costs of work, the stochastic process for wages, or indeed any of the further assumptions in section 2.1.2 or 2.1.3.[7]

Total within-period resources are allocated to consumption and leisure. If the optimum of this allocation problem yields a strictly positive number of hours work, the first order condition for within-period optimality is that the ratio of the marginal utility of leisure to that of consumption, that is the Marginal Rate of Substitution, equals the after tax real wage.

For our specification of preferences, for $l_{h,t} < H$, this equation will be:

$$w_{h,t} = \frac{u_{l_{h,t}}}{u_{c_{h,t}}} = \alpha_{h,t} \frac{l_{h,t}^{-\theta}}{c_{h,t}^{-\phi}} \tag{13}$$

[6] When a wage changes stochastically, the response of hours worked will partly be due to the Frisch intertemporal substitution motive, but will also be affected by the change in the marginal utility of wealth due to the particular wage realisation.

[7] If the cost of work is not fixed but related to the number of hours worked, then we would have to use the wage rates net of these costs to consider the MRS equilibrium condition.

This equilibrium equation can be used to compute static labour supply elasticities.

Differentiating the within-period budget constraint (equation 3 in assumption *A2*) and the MRS equation (13) with respect to wages, and solving, we obtain an expression for Marshallian elasticities for consumption and female leisure:[8]

$$\varepsilon_c^M = \frac{\partial \ln c}{\partial \ln w} = \frac{\theta w\left(H-l\right)+wl}{\theta c+\phi wl} \tag{14}$$

$$\varepsilon_l^M = \frac{\partial \ln l}{\partial \ln w} = \frac{\phi w\left(H-l\right)-c}{\theta c+\phi wl} \tag{15}$$

If preferences were Cobb-Douglas, θ and ϕ would both equal 1; and the Marshallian elasticities for consumption and for leisure would be equal to 1 and 0, respectively, if there were no unearned income or savings. In this case, if ϕ is greater than 1, the Marshallian elasticity for consumption is less than 1, and the leisure elasticity greater than 0. If preferences were a standard CES, the values of θ and ϕ would be equal. In section 6 below, we show how much heterogeneity is introduced through our more general specification in equations (14) and (15) and through allowing for unearned income.

We can obtain Hicksian elasticities from the Marshallian elasticities by using the Slustky equation. We first calculate the income elasticities by differentiating the MRS equation and the budget constraint with respect to income:

$$\varepsilon_c^y = \frac{\partial \ln c}{\partial \ln y} = \frac{\theta y}{\theta c+\phi wl} \tag{16}$$

$$\varepsilon_l^y = \frac{\partial \ln l}{\partial \ln y} = \frac{\phi y}{\theta c+\phi wl} \tag{17}$$

This then gives the expressions for the Hicksian elasticities:

$$\begin{aligned}
\varepsilon_c^H &= \varepsilon_c^M + \frac{\partial \ln c}{\partial \ln(c+wl)}\frac{wl}{(c+wl)} &&= \frac{-c}{\theta c+\phi wl}\\
\varepsilon_l^H &= \varepsilon_l^M - \frac{\partial \ln l}{\partial \ln(c+wl)}\frac{w(H-l)}{(c+wl)} &&= \frac{wz}{\theta c+\phi wl}
\end{aligned}$$

Several points are worth noting. First, despite their simplicity, these equations result in non-linear expressions for the elasticities that have the potential of varying greatly across consumers and do not aggregate in a straightforward way. Second, for the specification we have used, the Marshallian and Hicksian elasticities depend only on ϕ and θ (and on the values of earnings and consumption). In particular, they do not depend on the inter-temporal parameters or on whether consumption and

[8] Taking the derivative of the budget constraint and the MRS equation and stacking them gives a matrix equation that can be inverted to give the elasticities:

$$\begin{bmatrix} 1 & \frac{wl}{c} \\ \phi & -\theta \end{bmatrix}\begin{bmatrix} \frac{\partial \ln c}{\partial \ln w} \\ \frac{\partial \ln l}{\partial \ln w} \end{bmatrix} = \begin{bmatrix} \frac{w(H-l)}{c} \\ 1 \end{bmatrix}$$

leisure are separable in the utility function. Therefore, the specification of intertemporal utility in equation (5) is not needed to recover Marshallian and Hicksian elasticities. Third, by log-linearizing the Marginal Rate of Substitution equation (13), we can derive an expression that can be used to estimate the parameters needed to identify the Marshallian and Hicksian elasticities. Taking logs of equation (13), and noticing from equation (2) that $\log \alpha_{h,t} = \psi z_{h,t} + \chi_{h,t}$, we obtain:

$$\ln w_{h,t} = \phi \ln c_{h,t} - \theta \ln l_{h,t} + \psi_z z_{h,t} + \psi_0 + \chi_{h,t} \tag{18}$$

As we discuss in section 3 below, the first step of our empirical approach uses this equation to yields estimates of the parameters that enter $\alpha_{h,t}$ (that is the vector ψ), as well as ϕ and θ. These parameters pin down the within-period elasticities. In addition, as is well known, Frisch intertemporal elasticities, which we discuss below, must be at least as great as Hicks elasticities. Thus, static elasticities provide a bound on the intertemporal elasticity. In other words, we are able to obtain bounds on intertemporal elasticities without imposing any of the additional assumptions on the economic environment specified in section 2.1.2 or 2.1.3. This is particularly useful if there is limited data or complications in estimating Frisch elasticities directly.[9]

2.4 Euler equations

Having considered the intratemporal margin conditional on participation (MRS), we now characterise the intertemporal optimization conditions for the household which are given by a set of Euler equations. In addition to the assumptions of the previous section, we now make the additional assumptions specified in section 2.1.2. However, we can proceed in this subsection without invoking the assumptions of section 2.1.3 which specify the exact form of the income process, the uncertainty that individuals face and other details of the economic environment. We can bring the intertemporal optimization conditions to data without specifying completely what is in the household's information set. Therefore, our estimates of the intertemporal elasticities will be robust to how the economic environment is specified in section 2.1.3.

While in principle we could consider either the Euler equation for hours or that for consumption, only one is relevant, when coupled with the intratemporal condition (13). If we were to use the Euler equation for labour supply, we would need to consider corner solutions at different points in time (and the dynamic selection problems these involve). Instead, we focus on the Euler equation for consumption. Assumption *A5* rules out borrowing constraints. Therefore such a household is not at a corner solution for savings and, for them, the following intertemporal condition holds:

$$E\left[\beta\left(1+r_{t+1}\right)\frac{u_{c_{h,t+1}}(\cdot)}{u_{c_{h,t}}(\cdot)}\middle| I_{h,t}\right] = 1 \tag{19}$$

[9] In the context of quasilinear utility as used by Chetty (2012), the Frisch elasticity equals the Hicks elasticity (and the Marshallian) because there are no wealth effects on hours of work.

The term $I_{h,t}$ denotes the information available to the household at time t that conditions the expectations about the future.

The evolution of the marginal utility of consumption can then be written as:

$$\beta\left(1+r_{t+1}\right)u_{c_{h,t+1}}\left(\cdot\right)=u_{c_{h,t}}\left(\cdot\right)\epsilon_{h,t+1} \tag{20}$$

where $\epsilon_{h,t+1}$, whose conditional expectation is 1, represents the innovation to the expected discounted marginal utility of consumption and it incorporates innovations about present and future expected wages, male earnings and interest rates, as well as the demographic variables $z_{h,t+1}$ and taste shifters $\chi_{h,t+1}, \zeta_{h,t+1}$. The fact that equation (20) holds under different assumptions on the stochastic process is the strength of this approach.

The log of the marginal utility of consumption is given by:

$$\ln u_{c_{h,t}}=-\gamma\ln M_{h,t}-\phi\ln c_{h,t}+\varphi P_{h,t}+\pi z_{h,t}+\zeta_{h,t} \tag{21}$$

Taking the log of the Euler equation (20) and rearranging using equation (21) we get:[10]

$$\eta_{h,t+1}=\kappa_{h,t}+ln\beta+ln(1+r_{t+1})-\phi\Delta\ln c_{h,t+1} \tag{22}$$

$$-\gamma\Delta\ln(M_{h,t+1})+\varphi\Delta P_{h,t+1}+\pi\Delta z_{h,t+1} \tag{23}$$

where $\eta_{h,t+1}\equiv\ln\epsilon_{h,t+1}-E\left[\ln\epsilon_{h,t+1}\left|I_{h,t}\right.\right]+\Delta\zeta_{h,t+1}$ and $\kappa_{h,t}\equiv E\left[\ln\epsilon_{h,t+1}\left|I_{h,t}\right.\right]$. The error term, η, combines the expectation error and the taste shifters that are unobserved to the econometrician. Notice that $E\left[\eta_{h,t+1}\left|I_{h,t}\right.\right]=0$ by construction. We discuss the identification and estimation of the parameters of this equation in section 3.3 below. Frisch wage elasticities on the intensive margin can be calculated directly from the Euler equations and are given by the following expressions (derived in Appendix A):

$$\varepsilon_c^F=\frac{-\frac{u_{cl}}{u_l}w}{c\left(\frac{u_{ll}}{u_l}\frac{u_{cc}}{u_c}-\frac{u_{cl}^2}{u_c u_l}\right)}=\frac{w\gamma c^{-\phi}l}{\left\{\gamma\phi\alpha l^{1-\theta}+\theta\gamma c^{1-\phi}+M\phi\theta\right\}} \tag{24}$$

$$\varepsilon_l^F=\frac{\frac{u_{cc}}{u_l}w}{l\left(\frac{u_{ll}}{u_l}\frac{u_{cc}}{u_c}-\frac{u_{cl}^2}{u_c u_l}\right)}=\frac{-\left(\gamma c^{1-\phi}+M\phi\right)}{\left\{\gamma\phi\alpha l^{1-\theta}+\theta\gamma c^{1-\phi}+M\phi\theta\right\}} \tag{25}$$

The Frisch interest rate elasticities are given by

$$\varepsilon_c^{F_R}=\frac{\frac{u_{ll}}{u_l}-\frac{u_{cl}}{u_c}}{c\left(\frac{u_{ll}}{u_l}\frac{u_{cc}}{u_c}-\frac{u_{cl}^2}{u_c u_l}\right)}=\frac{-\theta M}{\left(\theta\phi M+\gamma\theta c^{1-\theta}+\gamma\theta\alpha l^{1-\theta}\right)} \tag{26}$$

$$\varepsilon_l^{F_R}=\frac{\frac{u_{cc}}{u_c}-\frac{u_{cl}}{u_l}}{l\left(\frac{u_{ll}}{u_l}\frac{u_{cc}}{u_c}-\frac{u_{cl}^2}{u_c u_l}\right)}=\frac{-\phi M}{\left(\theta\phi M+\gamma\theta c^{1-\theta}+\gamma\theta\alpha l^{1-\theta}\right)} \tag{27}$$

[10] We assume that the marginal utility of consumption and the discount factor are always strictly positive, and that the real interest rate r_{t+1} is bounded away from -1, so that the support of $\epsilon_{h,t+1}$ is $\Re^+$.

3 Empirical strategy

In this section, we discuss our empirical approach, the identification assumptions we make, and what type of variability in the data identifies which parameters. Our empirical strategy proceeds in three steps, with each successive step requiring stronger assumptions for identification. We use US household level data on consumption, labour supply, earnings and wages (as well as a variety of demographic variables) to estimate the model parameters.

We estimate a complete model of individual labour supply. Considering the whole model in its various components is essential if one wants to use it to evaluate the size and aggregation properties of the elasticity of labour supply to changes in wages. Each component provides an important element of the overall effect which, as we discuss below, affects our interpretation of the aggregate data and provides important insights on the relationship between 'micro' and 'macro' elasticities. From an empirical point of view, the approach we take is attractive because it makes, at each stage, the weakest possible assumptions and identifies sets of structural parameters on the basis of these assumptions.

We estimate all the components of the model sketched in the previous section by exploiting different sets of equilibrium conditions and different sources of variability in the data. In the first step, we consider only the static first order condition that determines within-period optimal allocations, conditional on participation. This first set of parameters can therefore be identified while being agnostic about intertemporal conditions and on life cycle prospects. We adjust our estimates for the fact that we consider an equilibrium condition that holds *conditional on participation*. These parameters, as discussed in section 2.3, can be used to derive expressions of the Marshallian and Hicksian elasticities that are of interest in their own right. Following the most recent literature on labour supply (such as Blundell et al. (1998)), we refrain from using variation in individual wages. Instead we exploit variation induced by changes in taxation and/or aggregate demand for labour and make use of changes in cohort and education groups' average wages over time.

In the second step, the parameters that govern the intertemporal allocation of resources are identified using the Euler equation for consumption. Of course, in this second step, we make use of an additional set of assumptions, specified in section 2.1.2. However, we can still identify these parameters without specifying the entire life cycle environment faced by the household (the assumptions in 2.1.3. For instance, we can be silent about pension arrangements or the specifics of the wage and earning processes we consider. Crucially in our context, we can be agnostic about the specification of fixed costs that make the extensive margin particularly important for characterising fully labour supply behaviour. To estimate the parameters of the Euler equation, we use only variability in intertemporal prices over time to avoid making strong assumptions (such as that of complete markets) on intertemporal trades.

Finally, in the third step, we characterise behaviour at the extensive margin, and this requires solving the entire model and, therefore, specifying completely the environment in which the consumer operates. This requires the set of assumptions in section 2.1.3. Obviously for our results on the size and variability of labour supply elasticities both at the intensive and extensive margins, some aspects of the environment would be more relevant than others. We identify the final set of parameters by calibration, that is matching a set of life cycle statistics.

We divide the discussion of our empirical strategy into three sections, corresponding to the three steps of the procedure: the Marginal Rate of Substitution, the Euler equation and the dynamic problem that determines the extensive margin. Before delving into that discussion, however, we briefly mention some econometric issues that are relevant for the estimation of the parameters that characterise the Marginal Rate of Substitution and the Euler equation. These issues are relevant when using equilibrium conditions and some assumptions about the nature of the random variables that enter the problem that can be expressed as a set of orthogonality conditions.

3.1 Using equilibrium conditions

When estimating the parameters that determine the MRS or those that enter the Euler equation, we make use of two sets of first order conditions to derive restrictions that can be imposed on the data to identify some of the structural parameters of our model. Although these sets of conditions are different, as one set is static in nature and one set is dynamic, they are of a similar nature, in that they can be reduced to an expression of the type

$$E[h(X;\theta)z] = 0 \tag{28}$$

where $h(.)$ is a function of data X and parameters, θ, and is linear in the vector of parameters. z is a vector of observable variables that will be assumed to be orthogonal to h. The nature of the instruments that deliver identification depends on the nature of the residual h and, as we discuss below, is different when we estimate the MRS conditions or the Euler equations. However, in both cases, we exploit a condition such as (28).

In equation (28), one needs to normalise one of the parameters to 1. In the context of the MRS in equation (18), for example, we have set the coefficient on $\ln w_{h,t}$ to 1, but we could have set the coefficient on $\ln l_{h,t}$, or that on $\ln c_{h,t}$ to be 1. A well-known issue with many estimators in this class is that in small samples they are not necessarily robust to the normalisation used. A number of alternative estimators that avoid this issue are available, ranging from LIML-type estimators, to the estimator discussed in Alonso-Borrego and Arellano (1999), to the iterated GMM proposed by Hansen et al. (1996). We use the estimator proposed by Fuller (1977) to estimate both our MRS and Euler equations. While this estimator is not completely normalisation free, it is much less sensitive to the choice of normalisation than estimators such as 2SLS and GMM.

An additional advantage of the Fuller estimator is that it is known to have better bias properties than estimators such as 2SLS, when instruments are relatively weak. In what follows, we test the strength of our instruments comparing the values of the Cragg-Donald test statistic to the relevant entries of the table supplied in Stock and Yogo (2005).[11] For the Fuller estimator that we employ, these critical values are typically lower than those for 2SLS, and, unlike 2SLS, they are decreasing in the number of instruments used. We report further details on the Fuller estimator in Appendix B.

3.2 Step 1: Intratemporal margins

In step 1, we estimate the parameters of within-period utility function: θ, ϕ and α. As mentioned above, for households not at a corner, that is where the woman works, the relevant intra-temporal equilibrium condition is given by equation (18), which we reproduce here for convenience:

$$\ln w_{h,t} = \phi \ln c_{h,t} - \theta \ln l_{h,t} + \psi z_{h,t} + \psi_0 + \chi_{h,t} \tag{18}$$

where the vector $\psi z_{h,t} + \psi_0 + \chi_{h,t} = \ln(\alpha_{h,t})$ and the vector $z_{h,t}$ includes a number of observable variables, such as demographic variables, that are assumed to shift systematically the weight on leisure $\alpha_{h,t}$ in the generalised CES function. Notice the importance of the unobserved heterogeneity term $\chi_{h,t}$ in this equation: in its absence we would have an equation with perfect fit that would obviously be rejected by the data and would imply the ad-hoc consideration of measurement error in the relevant variables.

The econometric estimation of the MRS equation poses two problems. First, the subset of households for whom the wife works and the MRS condition holds as an equality is not a random subset. This would therefore imply that the unobserved heterogeneity term $\chi_{h,t}$ would not average out to zero and would be correlated with the variables that enter equation (18). Second, even in the absence of participation issues and corner solutions, it is likely that individual wages (and consumption and leisure) will be correlated with the unobserved heterogeneity term, so that the use of OLS to estimate such an equation would result in biased estimates of the structural parameters ϕ and θ. We discuss these two issues in turn.

For participation, we first specify a reduced form equation for the extensive margin. We then use a Heckman-type selection correction approach to estimate the MRS equation (18) only on the households where the wife works and augmenting it with a polynomial in the estimated residuals of the participation equation. Non-parametric identification requires that some variables that enter the participation equation do not enter the specification for the MRS: these variables are male earnings and male employment status.[12]

[11] The Cragg-Donald statistic is usually used to provide a test of underidentification. Stock and Yogo (2005) propose using it as a test of instrument relevance as well.

[12] We assume that $\chi_{h,t} = \beta_0 + \beta_1 e_{h,t} + \beta_2 e_{h,t}^2 + \beta_3 e_{h,t}^3$ and then compute $E[e_{h,t}^s | e_{h,t} > -\Pi Z_{h,t}], s = 1, 2, 3$ where $e_{h,t}$ is the normally distributed residual from the participation equation and $\Pi Z_{h,t}$ are the determinants of participation.

The participation equation we use is consistent with our structural dynamic model we use in step 3, in that it models participation as a function of the state variables of the dynamic problem in equation (4). However, at this step, we do not solve the full model explicitly. Besides simplicity, this approach has the advantage of delivering consistent estimates of the parameters of the MRS equation even if some of the details of our full dynamic model were mis-specified, such as the specification of the innovation process. In other words, this approach does not require us to take a stand on either the dynamic aspects of the consumer problem or on the details of the stochastic processes faced by consumers. As it is based on two-stage budgeting, it only considers the *within-period* optimization. As discussed in section 2.1.1, however, two stage budgeting does rely on the assumption that hours of work have no impact on future wages.

The reduced form participation equation we use to correct for selection bias in estimating the parameters of the MRS is consistent with the life-cycle model we describe above in that its arguments are the same as the state variables that enter the policy function that would result from the solution of the full model. However, it is an approximation to such a policy function. The relevant issue is whether such an approximation yields a control function sufficiently flexible to avoid the selection bias in the estimation of the MRS parameters. In other words: if our life cycle model holds, would the MRS equation including the control function derived from such reduced form participation equation yield consistent estimates? An analytical answer to this question is not available. We therefore performed some Monte Carlo exercises, which we discuss in Section 5.4. In particular, we solve and simulate the entire model using the parameters we estimate. We then apply the approach we have just described to estimate the MRS and we show we can recover back the same parameters.

The second issue in the estimation of equation (18) is that our measures of wages, which is obtained by dividing earnings by hours, might be correlated with the residual term $\chi_{h,t}$. This correlation could be due either to measurement error in hours or earnings or to the possible correlation between taste for leisure and heterogeneity in productivity. To avoid these problems, we use an instrumental variable approach and exploit only part of the observed variability in wages to identify the parameters of interest.

Various papers have used differential changes in wages and hours across education groups as a means to identify labour supply elasticities; for example MaCurdy (1983) and Ziliak and Kniesner (1999) both use age-education interactions as instruments for wages and hours in their MRS/labour supply conditions. Similarly, Kimmel and Kniesner (1998) use education interacted with a quadratic time trend as an instrument for wages. However, one concern with using this source of variation in wages and hours is that individuals with different levels of education might have different preferences for leisure and consumption. Moreover, the composition of education groups has changed substantially over time: an issue which may be particularly important for women.[13] One would expect these compositional changes to also lead to changes in the mix of ability and preferences of workers within each education group over time - rendering education invalid as an instrument.

Our approach is to use as instruments the interaction of ten-year birth cohort and education dummies with a quintic time trend. Using fully interacted cohort-education and year dummies would be equivalent to taking averages within cells defined by year, education and cohort groups, and so use group level rather than individual level variability, as in Blundell et al. (1998). Given our sample size, however, we do not adopt this approach, as it would result in taking averages over relatively small cells and, therefore, getting very noisy estimates. Using very finely defined and small groups can introduce the very biases grouping is meant to avoid.[14] Our use of a quintic time trend rather than fully interacted time dummies, whilst in the spirit of Blundell et al. (1998), helps smooth intertemporal movements in the wages, consumption and hours for each of our cohort-education groups.

In our estimating equation, the z vector includes: log family size, race of woman, quartic in age of woman, an indicator for the presence of any child, number of children aged 0-2, number of children aged 3-15, number of children aged 16-17, the number of individuals in the household 65 or older, region dummies and season dummies, and, most importantly, cohort-education dummies. In other words, we allow all these variables to shift the taste for leisure through an effect on the term $\alpha_{h,t}$ in the CES utility function. A corollary of putting variables such as cohort and education dummies in the vector z is that we do not exploit the variation in wages (and leisure and consumption) over these dimensions to identify the structural parameters ϕ and θ. In our estimation, we also control for year dummies, therefore removing year to year fluctuations from the variability we use to identify the parameters of interest. The inclusion of year dummies, which is consistent with Blundell et al. (1998), can be justified because aggregate fluctuations change the selection rule in ways that are not fully captured by the Heckman selection model we considered.[15]

3.3 Step 2: Euler equation estimation

Step 2 uses the Euler equation (20) to estimate γ, which governs the intertemporal substitutability and non-separability between consumption and leisure. A natural approach to the estimation of equation (20) is non-linear GMM. However, given the nature of the data we have, it is only possible to bring is its log-linearized version, as in equation (23), to data. Moreover, as discussed in Attanasio and Low (2004), the small sample properties of non-linear GMM estimators can be poor when applied to Euler equations similar to that we are studying. We therefore focus on the estimation of equation (23).

[13] In 1980, 19.4% of married women had not attained a high school diploma, and only 18.4% had obtained a college degree in our data. By 2012, these proportions had changed to 9.7% and 36.5% respectively.

[14] Consider for example a case where both group average hours and group average wages are driven upwards by the inclusion of an individual with a particularly high wage and large unobserved tastes for work

[15] We have also run specifications where we do not control for time dummies in the MRS and checked that our results are not affected much by the introduction of the time dummies.

The identification and estimation of the parameters of this equation depends, obviously, on the nature of the 'residual' term $\eta_{h,t+1}$ on its right-hand-side. As noted above, $\eta_{h,t+1}$ contains expectations errors ($\varepsilon_{h,t+1}$) and taste shifters unobservable to the econometrician ($\zeta_{h,t+1}$). The expectation errors will be correlated with the realised values and so we would have biased parameter estimates if we used realised values directly. However, the rational expectations assumption that is typically invoked, implies that any variable known to the household at time t is a valid instrument. On the other hand, to achieve consistency using such an argument, it will be necessary to exploit explicitly the time series variation and, therefore, as discussed in Attanasio and Low (2004), a long time series is required to achieve consistency.[16]

Even if we can use a sample that covers a large number of time periods, we need to assume that the lagged variables that are used as instruments are uncorrelated with the innovations to the taste shifters $\Delta\zeta_{h,t+1}$. This is trivially true if individual taste shifters are constant over time or if they are random walks. In what follows we maintain that one of these two assumptions holds, which can be in part tested by considering over-identifying restriction tests.

The nature of the data we use, the Consumer Expenditure Survey (CEX), which we describe in section 4, poses some additional challenges to the identification and estimation of equation (23). In particular, although the CEX now covers many years over which we can consider quarterly data, as in many other household surveys, each household is only observed for a few quarters. Therefore, it is not possible to observe the same households over an extended period.

For this reason, we use a synthetic cohort approach (see Browning et al. (1985)). One can aggregate an equation such as (23) over certain groups with constant membership and then follow the average behaviour of the variables of interest (or their non-linear transformation) for such groups. A time series of cross sections can be used to construct consistent estimates of these aggregates and, in this fashion, use a long time period to estimate the parameters of the Euler equation and test its validity.

We define groups using ten year birth-cohorts. The assumption of constant membership of these groups might be questioned at the beginning and at the end of the life cycle for a variety of reasons, including differential rates of family formation, differential mortality and so on. To avoid these and other issues, we limit our sample to households whose husband is aged between 25 and 65 and where wives are aged between 25 and 60.[17]

Having identified groups and denoting them with the superscript g, we define as $X_{g,t}$ the (population) average for group g of the variable $X_{h,t}$. We then aggregate equation (23) across households

[16] The reason for the need of a long time series is that, even under rational expectations, expectations errors do not necessarily average out to zero (or are uncorrelated with available information) in the cross section, but only in the time series: expectation errors may be correlated with available information in the cross section in the presence of aggregate shocks. See the discussion in Hayashi (1987), Miller and Sieg (1997), Attanasio (1999), or Attanasio and Weber (2010).

[17] If credit constraints are binding, the Euler equation will not be holding as an equality. Very young consumers are excluded because they are more likely to be affected by this issue. For older consumers, in addition to changes in labour force participation and family composition, health status also changes in complex ways that maybe difficult to capture with the taste shifters that we have been considering.

belonging to group g to get:

$$\begin{aligned} \eta_{g,t+1} \quad = \quad & \kappa_{g,t} + ln\beta + ln(1 + r_{t+1}) - \phi \Delta \ln c_{g,t+1} + \\ & \Delta \ln \alpha_{g,t+1} - \gamma \Delta \ln(M_{g,t+1}) + \varphi \Delta P_{g,t+1} + \pi \Delta z_{g,t+1} \end{aligned} \tag{29}$$

For this approach to work, however, it is necessary that the relationship one studies is linear in parameters. If $M_{h,t}$ were observable, this would be the case for equation (29). However, $M_{h,t}$ is a non-linear function of data *and* unobserved parameters, so that, in principle it cannot be aggregated within groups to obtain $M_{h,t}$. To address this issue we use the fact that the parameters that determine $M_{h,t}$ can be consistently estimated, as discussed in section 3.2, using the MRS conditions. Given these, consistent estimates of the parameters that enter $M_{h,t}$, one can construct consistent estimates of $M_{h,t}$ and, effectively, treat it as data which can be aggregated across households to give $M_{g,t}$. We calculate $\alpha_{h,t}$ (needed to estimate $M_{h,t}$) for each individual household by evaluating $\exp(\psi_0 + \psi_z z_{h,t} + \chi_{h,t})$, where $\chi_{h,t}$ is the sum of the residuals, time dummies and Heckman selection terms from our MRS equation.[18] Once we have obtained $\alpha_{h,t}$ the calculation of $M_{h,t}$ is straightforward.

Finally, the quantities that enter equation (29) are population means of the relevant variables and, as such, are not directly observable. However, we can obtain consistent estimates of these quantities from the time series of cross sections that we have. We can therefore substitute these observable quantities and obtain:

$$\begin{aligned} \widetilde{\eta}_{g,t+1} \quad = \quad & \overline{\kappa} + ln\beta + ln(1 + r_{t+1}) - \phi \Delta \overline{\ln c_{g,t+1}} + \\ & - \gamma \Delta \ln(\overline{\widehat{M}_{g,t+1}}) + \varphi \Delta \overline{P_{g,t+1}} + \pi \Delta \overline{z_{g,t+1}} \end{aligned} \tag{30}$$

The residual term $\widetilde{\eta}_{g,t+1}$ now includes, in addition to the average of the expectation errors and of the changes in taste shifters, several other terms. In particular, it includes: (i) a linear combination of the difference between the population and sample averages at time t and $t+1$ for all the relevant variables (induced by the fact that we are considering sample means rather than population means for group g); (ii) the difference between the (consistently) estimated $M_{g,t}$ and its actual value (induced by estimation error in the parameters of the MRS); (iii) the difference between the innovation over time to the average value of $\kappa_{g,t}$, which we have denoted with the constant $\overline{\kappa}$.

All the variables on the right hand side of equation (30) are observable. We can therefore use this equation to estimate the parameters of interest. However, the instruments need to be uncorrelated with

[18]This must also be calculated for non-participants for whom we do not have estimates of the MRS residuals. We do this by imputing wages to those out of work using a regression of wages on family characteristics and instrument set, calculating a lower bound on what this would imply for their residuals given our MRS coefficients and their non-participation, and then adjusting these residuals such that for all participants and non-participants $E[v_{h,t}] = 0$.

$\widetilde{\eta}_{g,t+1}$.[19] The covariance structure of the $\widetilde{\eta}_{g,t+1}$ is quite complex. The contemporaneous covariance of $\widetilde{\eta}_{g_i,t+1}$ and $\widetilde{\eta}_{g_j,t+1}$ is not, in general zero, as aggregate shocks will have effects that correlate across the various groups. When computing the variance-covariance matrix of the estimates, this structure should be taken into account. Whilst it is in principle possible, given our assumptions, to estimate the variance covariance of $\widetilde{\eta}_{g,t+1}$ from the estimated parameters, in practice it turns out to be cumbersome, as there is no guarantee that, in small samples, these estimates are positive definite. Given these difficulties, we decided to follow a different and, as far as we know, novel approach based on bootstrapping our sample, with a structure consistent with the basic assumption of our model. We describe the bootstrapping procedure in detail in Appendix D.

In principle, steps one and two of the estimation could be followed without making parametric assumptions about the form of utility and instead estimating demand directly. However, this would require that the demand functions satisfy integrability conditions. Further, the actual underlying utility function would still need to be recovered to use in the third step of the estimation when carrying out the explicit solution to the utility maximisation problem. We follow the alternative approach of making functional form assumptions and allowing for a general and flexible parametric specification of utility.

3.4 Extensive margins

Step 3 estimates the remaining parameters of the model, including the fixed costs of work and child-care costs which drive the extensive margin decision. When considering this extensive margin, it is necessary to solve explicitly the dynamic problem. This involves the third block of assumptions in section 2.1.3 that specify completely the economic environment the individual households live in, including both present and future conditions. We solve the model numerically and use the numerical solution to estimate and calibrate the model parameters.[20]

We take as given the estimates of the other parameters that we obtained from the MRS and the Euler Equation, and obtain some parameters from other sources: either the literature or auxiliary regressions. Armed with these parameters, we use a number of life cycle facts and match similar moments computed by simulating our model to obtain the remaining parameters. In particular, we chose to calibrate the model so that it would match key aspects of the extensive margin: the participation rate, the participation rate of mothers and average wage growth of participants (which is endogenous because of selection). We will then simulate the model for a large number of individuals to study the properties of individual and 'aggregate' labour supply.

[19] As noted by Deaton (1985) and discussed extensively in the context of the CEX by Attanasio and Weber (1995), the use of sample rather than population averages for all the 'group' variables induces an MA(1) in the residuals, induced by the sampling variation in the rotating panel structure. We need to assume that the instruments are not correlated with the (average) estimation error of the $M_{h,t}$ or with the innovations to the higher moments of the expectation errors $(\kappa_{g,t} - \overline{\kappa})$. This last assumption is discussed in Attanasio and Low (2004).

[20] See Appendix G for details of the numerical solution.

Goodness of fit Having estimated all parameters, we simulate the model and check whether it is able to fit several features of the data, over and above those that have been used to derive the parameter estimates (either by econometric methods or by calibration). In particular we explore: participation and hours life cycle profiles, participation rates conditioning on several characteristics such as motherhood, and the distribution of hours worked and of wages.

4 Data and descriptive statistics

Except for information on interest rates, which is defined as the return on 3 month Treasury Bills, taken from the Federal Reserve Bank of St Louis, and the price level, we take our data from the American Consumer Expenditure Survey (CEX) for the years 1980-2012. In the CEX, households are interviewed up to five times, but only very limited baseline questions are asked in the first interview. In the second to fifth interviews households are asked detailed recall questions on expenditures as well as on the demographics, incomes and labour supply of household members.

We calculate gross hourly wages for individuals using information on the value of each individual's last pay cheque, the number of weeks it covered and the typical number of hours worked per week over the previous year. Net wages are then calculated by subtracting marginal federal income tax rates generated using the NBER TAXSIM model (Feenberg and Coutts, 1993).[21] We deflate all expenditures, wages and incomes using the Consumer Price Index for the appropriate period. Weekly leisure is calculated by subtracting weekly hours worked from the maximum number an individual has to divide between leisure and labour supply per week (which we set to 100). Our definition of consumption covers nondurable goods excluding medical and education spending. We divide quarterly consumption spending by 13 to put it in weekly terms.

Our sample consists of couples where the female is aged between 25 and 60 and males are aged between 25 and 65. We drop those in rural areas and those in the top 1% of the distribution of consumption and net wages. We also trim those who are seen to earn less than 3 quarters of the national minimum wage in any given year, and those who are employed but who report working less than 5 hours a week. Since labour supply and income questions are (almost always) only asked in the second and fifth interviews, we drop responses from interviews apart from these two. Our sampling choices leaves us with a sample of just under 79,000 households (50,895 where the female is working).

4.1 Statistics on wages and hours

Before presenting our structural estimates, it is useful to consider some of the main features of the data. Married women have seen large changes in their wages, hours and patterns of employment over our sample period, as Table 1 illustrates using data from three particular years (1980, 1995 and 2012).

[21] We are grateful to Lorenz Kueng for making his mapping of the CEX to TAXSIM publically available.

Table 1: Descriptive statistics for married women, 1980, 1995 and 2012

	1980	1995	2012
% Employed	60.0	69.8	61.9
% Workers part-time	28.4	23.7	20.6
No. of children	1.25	1.15	1.17
Education			
% Less than high school	19.4	12.3	9.7
% High school	44.1	36.8	25.3
% Some college	18.1	25.3	28.5
% Degree or higher	18.4	25.5	36.5
Hours (workers)			
All	35.2	37.5	38.4
Less than high school	34.9	37.4	34.2
High school	35.2	36.2	38.6
Some college	35.0	36.7	37.1
Degree or higher	35.5	39.7	39.5
Part-time	21.5	22.0	22.7
Full-time	40.6	42.3	42.4
Hourly net wages ($ 2016)			
All	15.58	16.63	18.95
Less than high school	12.16	11.23	11.33
High school	14.22	13.41	14.61
Some college	16.62	16.41	17.28
Degree or higher	19.30	22.26	23.20
Sample sizes			
All	2,199	2,064	2,026
Workers	1,318	1,441	1,254

Part-time is defined as working less than 35 hours per week.

Employment rates increased from 60% in 1980 to 69.8% in 1995 before falling back to 61.9% in 2012.

While employment fluctuated over the business cycle, average hours worked among workers rose steadily from 35.2 hours per week in 1980 to 37.5 in 1995 to 38.4 hours in 2012. This increase reflects two developments. The first is a steady decrease in the proportion of employed women who were working part-time (defined here as working fewer than 35 hours per week), which fell from 28.4% in 1980 to 20.6% in 2012. The second is an increase in the hours worked by both part-time and full-time workers. Full-time increased their average hours from 40.6 per week in 1980 to 42.4 hours in 2012. Over the same period, part-time workers increased their average hours from 21.5 to 22.7 per week.

Table 1 also shows wage changes between the three years. Unsurprisingly real wages increased over this period from \$15.58 an hour in 1980 to \$16.63 in 1995 to \$18.95 in 2012. There were however differences in the rate of increase by education group. The wages of those with less than high school education actually fell slightly from \$12.16 in 1980 to \$11.33 in 2012. By contrast, married women with a college degree or higher saw a 20% increase in their wages between 1980 and 2012 (from \$19.30 to \$23.20). This increase in the education premium has been attributed to skill-biased technological change which outstripped the supply of educated workers (Goldin and Katz, 2007).

Changes in hours worked across education groups appear to mirror these patterns. While all education groups worked very similar hours in 1980, by 2012 those with a college degree were working on average five hours more per week than those with less than high school education, although the fraction with a college degree has markedly increased over the period.

4.2 Cohort averages

In what follows, we exploit differences in the rate of growth of wages and hours over time across education-cohort groups to identify the relevant elasticities. As the CEX is not a panel data, rather than following individuals over time, we follow group averages. In particular, we separate households into birth cohorts and examine changes in wages and hours by education within each cohort group. The advantage of considering the variability over time of a given cohort, is that their composition is unlikely to change as it is relatively rare for workers to increase their educational qualifications after age 25. This is the approach taken in Blundell et al. (1998), and the approach we adopt to estimate our MRS equation.

Once we look within cohorts (in the bottom panel of Figure 1), the differences in the evolution of hours worked for workers with more than a high school education and those with less are clearly much smaller than they appear in Table 1. Taking the 1950s cohort the net wages of those with more than high school education increased from an average of \$16.90 per hour in 1980 to \$21.40 in 2012 (an increase of 27%), while the wages of those with less than high school education only increased by 19% from \$13.40 to \$16.00. Despite this, the average weekly hours of less educated worked actually increased by more than those from the more educated group (increasing from 36.8 hours per week to 38.2 compared to an increase from 37.4 to 38.5 for those with more than high school education).

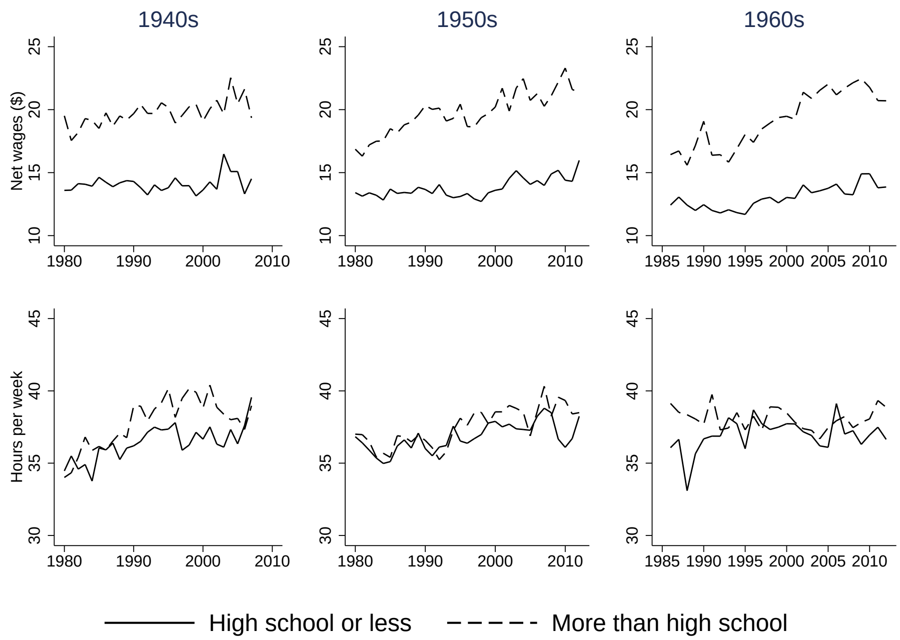


Figure 1: Wages and hours by education group and cohort

4.3 Individual variation in hours and wages

In addition to these changes in average hours and wages over our sample period, there are two important issues at the individual level: first, what is the relative importance of the intensive and extensive margins in the raw data; second, what fraction of individuals are experiencing changes in hours or wages over time.

We think of the individual extensive margin decision as being whether to incur a fixed cost $F(a_{h,t})$ and participate in the current quarter, and the intensive margin decision as being over how many hours to work per week (when working). An additional labour supply response may be through changing weeks worked per quarter. However, we are not able to esitmate this margin of adjustment because of data limitations: the CEX asks current workers about the number of weeks they worked over the previous year rather than the previous quarter.

If workers do partially adjust hours worked within a quarter by changing the number of weeks worked, rather than just their weekly hours conditional on participation, then our approach will miss some relevant variation in workers' labour supply choices. The key question is how much of the variance of workers' quarterly hours could conceivably be driven by differences in weeks worked within a quarter rather than hours per week. Table 2 decomposes the variance of log annual hours into that due to variation in log annual weeks, log workers' typical weekly hours, and their covariance. The first panel shows this breakdown for the entire sample of workers. It is clear that the variance in annual weeks worked makes for the larger share (around two thirds) of the total variance in hours worked. However, much of this could be due to some workers not participating for entire quarters: our extensive margin. In the second panel, we therefore restrict our sample to workers who we observe working for more than 39 weeks (and thus could not have been unemployed for a complete quarter). Focusing attention on these workers, who account for 84% of the total, we see that almost all of the variance in annual hours is a result of differences in hours worked per week, with differences in weeks worked having a negligible contribution. In the third panel, we restrict our sample further to those working exactly 52 weeks per year and notice that even among workers who do not differ in the number of weeks worked, the variance in log hours per week remains substantial (at 0.08).

These results suggest that among participating workers, hours worked per week is the key margin by which workers adjust their quarterly hours, lending support to our approach of modelling weekly hours choices to get at workers' intensive margin decisions. In Appendix B, we check the robustness of this strategy by examining how our estimates and results are affected when we replace our current measure of hours worked with a measure of annual hours worked. It turns out that the elasticities and parameter estimates under this approach are similar to the results from our main specification.

Table 2: Variances of log hours per and week and log annual weeks worked, 2012

	Less than high school	High school	Some college	Degree or higher	All
All workers					
Variance (ln hours per week)	0.148	0.117	0.128	0.126	0.126
Variance (ln weeks per year)	0.550	0.271	0.231	0.482	0.367
Covariance (ln hours, ln weeks)	0.031	0.046	0.010	0.028	0.027
Variance (ln annual hours)	0.761	0.479	0.380	0.665	0.546
Working at least 39 weeks (84% of workers)					
Variance (ln hours per week)	0.061	0.040	0.086	0.110	0.086
Variance (ln weeks per year)	0.001	0.003	0.003	0.005	0.004
Covariance (ln hours, ln weeks)	-0.001	0.001	0.002	0.000	0.001
Variance (ln annual hours)	0.062	0.042	0.090	0.115	0.090
Working 52 weeks (69% of workers)					
Variance (ln hours per week)	0.064	0.031	0.068	0.117	0.080

A further question is whether individual workers adjust their weekly hours at all in response to wage changes, or whether there are market frictions which prevent this. To understand how much individuals vary their weekly hours, Table 3 shows the proportion of workers who changed their hours from the second CEX interview to the fifth interview (a period of nine months). While it is true that most women do not change their hours within this period, a substantial fraction (46%) do. Around a quarter of individuals change their weekly hours by 1-5 hours, and 2% change their hours by more than 20 hours. One question is whether these changes in hours are associated with new jobs. We do not observe whether individuals switch employers in the CEX, but do see if individuals enter a different code for their occupation. Those who change occupations show very similar hours changes to those who do not. A second question is whether these intensive margin changes in hours reflect changes in wages. The last two rows of Table 3 shows that those who saw larger real wage changes were more likely to change their hours: 47.5% of those who saw wage changes of over 5% did not change their hours, compared to 75.9% who experienced a wage change of 5% or less.

Table 3: Changes in Weekly Hours among the Employed

Change in Weekly Hours	No change	1-5 hrs	6-10 hrs	11-20 hrs	>20 hrs
All Workers	53.8%	25.2%	11.9%	6.9%	2.2%
Changes occ.	51.1%	25.6%	12.4%	7.6%	3.4%
No change in occ.	54.9%	24.9%	11.8%	6.6%	1.7%
Extent of Change in wages:					
< 5% wage change	75.9%	17.5%	4.6%	2.3%	0.71%
> 5% wage change	47.5%	27.5%	14.0%	8.2%	2.7 %

Notes: Changes in hours are measured between the 2nd and 5th interviews for individuals who are employed at each interview.

5 Results: Parameter Estimates, Calibration, Goodness of fit

In this section, we report the estimates of the structural parameters of our model we obtain from the three steps of the estimation procedure. In sections 5.1 and 5.2 we report the estimation results obtained using the MRS conditions and the Euler equation. In section 5.3, we discuss the calibration of the remaining parameters of the model that govern choices at the extensive margin. In the last subsection, we show how well the complete model fits a number of features of the data that were not explicitly used to obtain the parameter estimates.

5.1 MRS estimates

In Table 4, we report the estimates of key parameters for the MRS equation and tests on the quality of our instruments. The results for the participation model are reported in Appendix C.

The Cragg-Donald statistic for weak instruments in our MRS equation takes a value of 2.00 for 138 instruments, well above the relevant Stock and Yogo (2005) critical level of 1.69, and therefore suggesting that weak instruments are not a problem.[22] The Sargan test does not reject the null of no violation of the overidentifying restrictions.

Table 4: Estimation of MRS equation

Parameter	Estimate	(Standard Error)	[95% Confidence Interval]
θ	1.75**	(1.230)	[0.34,5.12]
ϕ	0.76***	(0.103)	[0.55,0.95]
Ψ			
$\ln(famsize)$	-0.32***	(0.037)	[-0.38,-0.23]
Has kids	0.07***	(0.021)	[0.04, 0.10]
No. of kids 0-2	0.15***	(0.030)	[0.10, 0.22]
No. of kids 3-15	0.06***	(0.017)	[0.04, 0.10]
No. of kids 16-17	-0.02**	(0.011)	[-0.05,0.00]
Heckman selection terms			
e_1	0.07	(0.167)	[-0.18, 0.48]
e_2	0.05	(0.172)	[-0.21, 0.51]
e_3	0.01	(0.052)	[-0.08, 0.13]
Joint test of selection terms (p-value)		0.87	
Cragg-Donald statistic		2.00	
Sargan statistic (p-value)		127.8 (0.66)	

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Additional controls for the number of elderly (aged over 65) individuals in the household, a quadratic in age, race, region, season, cohort-education interactions and year dummies. Consumption and leisure are instrumented with the interaction of cohort and education groups and a fifth-order polynomial time trend. Confidence intervals are bootstrapped with 1000 replications allowing for clustering at the individual level.

We estimate a value for θ of 1.75 and a value for ϕ of 0.76: there is much more curvature in utility on leisure than on consumption. A standard Cobb-Douglas specification imposes that $\phi = \theta = 1$, while

[22]The value of 1.69 is given for two endogenous variables and 100 instruments, and given that the critical values for a maximum 5% relative bias for the Fuller estimator are decreasing in the number of instruments, the use of this test statistic is conservative.

a standard CES specification imposes that $\phi = \theta$. The values of θ and ϕ we estimate are however quite different. We test the restrictions implied by Cobb-Douglas and standard CES and specifications in our framework using a wild-cluster residual bootstrap. The Cobb-Douglas specification for preferences is rejected at the 5% level (p-value 0.01), while the standard CES specification is rejected at the 10% level (p-value 0.06).

Table 4 also shows the coefficients attached to some of the variables included in $z_{h,t}$ (a dummy for having children, the number of children of various ages and for family size), reflecting the impact of these demographic variables on the MRS.[23] A positive coefficient on one of these variables implies that women will supply less hours of work in the market, for a given level of consumption and wages, when this variable increases: a larger value for ψ means, other things equal, a higher marginal utility of leisure. The positive and significant coefficient on the dummy for having children thus indicates that the presence of children tends to reduce labour supply, but the effect of children on hours worked depends on the age of the children. The coefficient on the number of children aged 0-2 is positive and highly significant, on children aged 3-15 the coefficient is positive, but smaller; for older children, the coefficient is negative.

We include three Heckman selection terms corresponding to the first, second and third moments of the truncated normal distribution (as described in footnote 12). These terms are not significant, individually or jointly, with a test of whether they enter the MRS equation giving a p-value of 0.87. One possible reason for this is that our other controls (including time dummies) already correct for the biases resulting from selective participation.

5.2 Euler equation estimates

Table 5 shows the results we obtain from estimating the Euler equation (30). Our instruments are second, third and fourth lags of $\ln M_{g,t}$ and the lagged real interest rate, and we have two endogenous variables $\phi(\Delta \ln c_{g,t} + \ln(1 + r_{t+1}))$ and $\Delta \ln M_{g,t}$. We place the second of these on the left-hand side of the equation. With only one left-hand side endogenous variable, the Cragg-Donald test for weak instruments is equivalent to a standard F-test of the instruments' joint significance in the first stage regression. The critical values of these F-tests suggest that the instruments are highly correlated with the dependent variable (with an F-statistic of 7.95) but less strongly correlated with our choice of left-hand side variable (with an F-statistic of 2.08). The relevant Stock and Yogo test statistic for having less than a 5% relative bias in our parameter estimates when there are four instruments and one left-hand side endogenous variable is 7.63. When we carry out a Sargan test for the Euler equation, we fail to reject the null of over-identification (p-value 0.13) as we do for the MRS.

We estimate a value for γ of 2.07. This value is significantly different from zero at the 10%

[23] A complete set of parameter estimates is in the Appendix.

Table 5: Estimation of Euler equation

Parameter	Estimate	(Standard Error)	[95% Confidence Interval]
γ	2.07*	(0.656)	[-0.11,2.60]
$\bar{\kappa} + \ln(\beta)$	0.03	(0.040)	[-0.08, 0.10]
π			
$\ln(famsize)$	-0.47	(0.244)	[-0.69, 0.31]
Has kids	0.05	(0.069)	[-0.09, 0.19]
No. of kids aged 0-2	0.22	(0.099)	[-0.05, 0.35]
No. of kids aged 3-15	0.03	(0.038)	[-0.06, 0.09]
No. of kids aged 16-17	0.03	(0.071)	[-0.11, 0.18]
First Stage F-stats (p-values)			
$-\phi(\Delta \ln c_{g,t} + \ln(1 + R_{t+1})$		7.95 (<0.001)	
$\Delta \ln M_{g,t}$		2.08 (0.08)	
Sargan statistic (p-value)		5.70 (0.13)	

N = 1,519. *p<0.10, ** p<0.05, *** p<0.01. Additional controls for season dummies, a quartic in age, the change in the proportion of households in each of four education groups, the change in proportion who are white, and the change in the average number of elderly individuals per household. Instruments are second, third and fourth lags of ln $M_{g,t}$, as well as the lagged real interest rate. Confidence intervals are bootstrapped with 1000 replications.

level, providing evidence that preferences are non-separable and that consumption and leisure are substitutes (as a a value of 0 for γ would imply additive separability in preferences over consumption and leisure). Since ϕ, θ and γ are all positive, the concavity requirements of the utility function are satisfied. The coefficients on the control variables included in the vector z_t are not significant which means demographics have no role over and above their impact on the relative weight on leisure within-period. We report results below where we impose that ξ, the parameter on participation in equation (5), is zero.[24]

5.3 Calibration of the remaining parameters

As discussed in section 2.1.3 and in section 3.4, to estimate the responsiveness on the extensive margin, we need to specify all the details of the model and to quantify each of its elements. There are three sets of parameters used in the calibration: those estimated via the MRS conditions and the Euler equation, those coming from external sources and those that we calibrate using the full model.

[24]When we include this term (instrumented with its own lags), it enters negatively and is highly insignificant.

To obtain the calibrated parameters of our model, we have to specify the ex-ante heterogeneity that is captured by z, χ and ζ in our estimation. To to this, we focus on the cohort of women born in the 1950s. We further assume there are nine different types of women within this cohort: one group of women who remain childless for the whole of their lifetime, and eight groups of women who differ by maternity experience. These women exogenously receive two kids but differ in the age at which the first child arrives. To determine when these children are born, we draw on Rendall et al. (2010) who use population and survey data sources to calculate the distribution of maternity age at arrival of the first child for different cohorts of women in various countries. Consistent with the distribution they provide for our cohort of interest, we assume 16% of women are childless, 27% have their first child at the age of 19, 12% at the age of 22, 11% at the age of 24, 5% at the ages of 26, 28, 30 and 32 and, finally, 14% at the age of 34. We assume that the second child arrives 2 years after the first.

External Parameters. Table 6 reports the estimated and external parameters used in the calibration. The first panel reports the estimated parameters from Tables 4 and 5 above. The second panel reports parameters which come from external sources.

Table 6: External Parameters

Estimated Parameters (from first-order conditions)		
Curvature on leisure	θ	1.75
Curvature on consumption	ϕ	0.76
Curvature on utility	γ	2.07
Exogenous Parameters		
Interest Rate (annual)	r	0.015
Regression Log Wage on Age and Age^2 (Men)	ι_1^m, ι_2^m	0.0684, -0.00065
Husband and Wife Wage Correlation	ρ	0.25
Standard Deviation of Permanent Shock (Men)	σ_{ξ^m}	0.077
Standard Deviation of Initial Wage (Men)	$\sigma_{\xi^m,0}$	0.54
Length of Life (in years)	T	50
Length of Working Life (in years)	R	40

We fix the annualized interest rate to equal the average real return on three monthly T-bill at 0.015. The deterministic component of the male earnings process is estimated from the CEX: we take the two parameters of a regression of husband log earnings on age and age squared. Both the innovations to male earnings and those to female wages are assumed to have a unit root, consistent with the evidence on men produced by MaCurdy (1983) and Abowd and Card (1989). This is in addition to heterogeneity in the initial value of wages.

The standard deviation of the innovation for husband's earnings is set to be 0.077, consistent with Huggett et al. (2011). Further, we estimate an initial standard deviation of husband earnings of 0.54. There is limited evidence on the variability of female wages and/or earnings. Further, in contrast with men, this statistic is highly affected by non-random self-selection into the labour market, therefore we calibrate the parameters that characterise the women's wage process within the model as explained below. Finally, we assume that the correlation coefficient between the two shocks (for husband and wife) is equal to 0.25 as estimated by Hyslop (2001).

As in Attanasio et al. (2008), there are two components to child care costs: the function $G(a_{h,t})$ and the price p. We estimate the function $G(a_{h,t})$ directly from data. In particular, for households where the mother is working, we regress total childcare expenditure on the age of the youngest child, the age of the oldest child, the number of children and a dummy that equals one if the youngest child is 0. The shape $G(a_{h,t})$ can be derived from the coefficients of this regression function, considering that in our model all women with children have two of them and at the same interval between children of two years.[25] This implies that the child care cost can be expressed as a function of the age of the oldest child.

Finally, we assume individuals in this cohort live for 50 years from age 22, with the last 10 in retirement, and we assume that the household receives a pension equal to 70% of the husband's earnings in the final working period.

Calibrated parameters. There are nine parameters that we calibrate within our decision model: the fixed cost of working, $\bar{F}$; the price of child care, p; the offered wage gender gap, y_0^f/y_0^m; the standard deviation of the permanent shock to women, σ_{ξ^f}; the standard deviation of the initial wage for women, $\sigma_{\xi^f,0}$; two parameters that determine exogenous wage growth, ι_1^f and ι_2^f; and the base weighting on leisure in the CES utility function, ψ_0, which, together with a set of demographics z and the estimates of ψ_z, determine the total weight on leisure in the utility function. Finally, we also calibrate the discount rate β.

The targets for the calibration are taken from the data for the cohort of women born in the 1950s and observed between age 25 and 55. We target the female participation rate, the participation rate of mothers, average hours worked, the observed wage gender gap, the observed variance wage growth, the observed initial variance in wages and the observed wage growth at two different stages of the life cycle. Finally, we target median wealth to median household income ratio as in Low (2005).

[25] Our estimate of $G\left(a_{h,t}\right)$ combines the cost of the first born child along with any subsequent costs associated with additional children who are born later. In this way, any economies of scale in child costs will be captured by $G\left(a_{h,t}\right)$, but we do not identify separately the marginal cost of extra children.

In Table 7, we report the value of the parameters we obtain in this calibration exercise as well as the value of the targeted moments in the data and in the simulated data. The monetary fixed cost of working is about 6% of median earnings of women aged 25 to 55. The monetary fixed childcare cost is up to 13% of median earnings of women aged 25 to 55 for a child age 0-2.

The initial offered wage gender gap that it is needed to target the observed wage gender gap of 0.72 is 0.74. Note however than in addition to the initial wage gender gap there is a further, exogenous wage gap that opens up through differential wage growth for men and women over the life cycle. In particular the calibrated exogenous wage growth implies that male wages are on average 77% higher by the age of 45 than at the moment of entering the labour market. In contrast for a female the figure will be only 31%. We calibrate the standard deviation of female wages innovations to be 0.063 and the standard deviation of the initial women's wages to 0.50.

5.4 Goodness of fit

Our next step is to show to what extent the model can account for some observed features of female labour supply behaviour that were not explicitly targeted in the calibration. The calibration was focused on averages taken over the life cycle. Our focus here is on life cycle paths and on the distribution of hours and wages. Figure 2 shows life cycle profiles in the simulations and in the data while Table 8 reports additional moments on heterogeneity.

Figure 2: Life Cycle Profiles: Baseline Model (solid black line) versus Data (dashed red line)

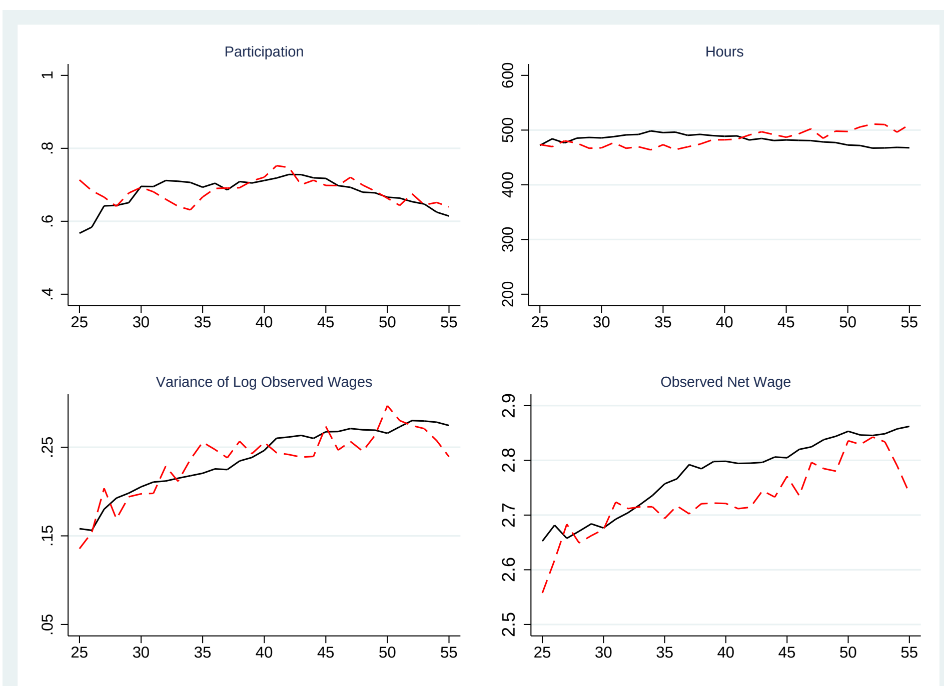


Table 7: Baseline economy: Calibrated Parameters and Targets

Parameters		**Value**
Constant term weight of leisure	ψ_0	4.20
Childcare Cost	p	967
Fixed Cost of Working	$\bar{F}$	468
Offered Wage Gender Gap at age 22	y_0^f/y_0^m	0.74
Standard Deviation of Permanent Shock (Women)	σ_{ξ^f}	0.063
Standard Deviation of Initial Wage (Women)	$\sigma_{\xi^f,0}$	0.50
Exogenous growth in offered wage	ι_1^f	0.052
Exogenous growth in offered wage	ι_2^f	-0.0006
Discount Factor (annualized)	β	0.99

Targets	**Data**	**Model**
Weekly hours worked	37.3	37.3
Participation Rate	0.677	0.678
Participation Rate of Mothers	0.538	0.546
Observed Wage Gender Gap	0.720	0.727
Observed Variance Wage Growth (Women)	0.005	0.005
Observed Initial Variance of Wages (Women)	0.15	0.15
Wage Growth (if younger than 40)	0.012	0.010
Wage Growth (if older than 40)	0.001	0.004
Median wealth to income ratio	1.84	1.80

Statistics for women born in the 1950s and aged 25 to 55. Wage growth is annual.

The life cycle path of female labour supply both at the extensive and intensive margin is similar in the model and in the data. From Table 8, the model does a good job in terms of participation of other demographic groups such as women who have no dependent children, and mothers of children aged 3 to 17. Regarding the intensive margin the distribution of hours worked is very close to the data, although the fraction of women working an average of 40 hours a week is higher in the data. Observed female wages and the variance of wages are increasing with age in our simulations, consistent with what we observe in the data. The profiles shown are shaped not only by our assumptions on the wage process, but also by the endogenous selection of women into the labour market. The distribution of observed wages is also similar to that in the data. It is worth emphasizing here that these facts about the comparison of wages in the model and in the data speak well about our model of selection into participation.

Table 8: Statistics on Hetereogeneity

	Data	Model
Participation Rate: Mothers with Children Aged 3-17	0.682	0.687
Participation Rate: Women without Dependent Children	0.730	0.694
Average Hours Worked 10th Percentile	20	25
Average Hours Worked 25th Percentile	35	31
Average Hours Worked 50th Percentile	40	38
Average Hours Worked 75th Percentile	40	44
Average Hours Worked 90th Percentile	48	48
Wage 10th Percentile	8.17	8.36
Wage 50th Percentile	15.09	16.02
Wage 90th Percentile	29.45	31.02

Women without dependent children are women who have never had children and those whose children are over 17.

Finally, we use a simulated sample to reestimate the MRS equation, employing the same procedure used in getting our estimates from the data and described in section 2.3. The estimates of the MRS parameters θ and ϕ that we obtained from actual data (and that were used to generate the simulated data) are almost identical to those we recover from the simulated data. Given the complexity of the model that includes discrete choices over the life cycle, this is an important validation of our strategy.

6 Labour supply elasticities

This section provides the key results of the paper. We use the estimates of the model to discuss implications for various wage elasticities. We start our discussion with the Marshallian and Hicksian elasticities that can be obtained from the MRS parameters estimated in step 1. We then move on to the Frisch elasticities at the intensive margin using step 2 estimates. We then simulate the full callibrated model to obtain elasticities at the extensive margin. In the final subsection, we look at aggregation issues and discuss the implications of our estimates for aggregate labour supply elasticities.

6.1 Marshallian and Hicksian hours elasticities

The first two columns in Table 9 show how the MRS parameters translate into within-period Marshallian and Hicksian wage elasticities for hours of work, for lesiure and for consumption. These elasticities vary according to family characteristics and the levels of consumption and leisure. We report elasticities at different points of the distribution of Marshallian elasticities to highlight the heterogeneity across individuals.

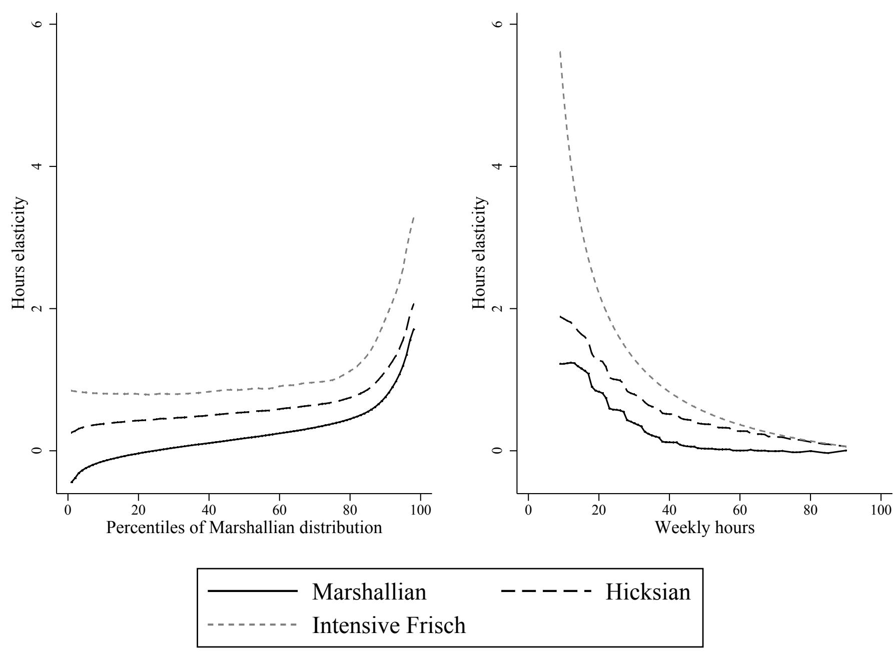
The median Marshallian hours elasticity is estimated to be 0.18, implying an upward sloping labour supply function. Given that the estimates of the utility function parameters satisfy quasi-concavity, Hicksian elasticities must be greater than Marshallian elasticities: for the household with the median Marshallian elasticity, the Hicksian hours elasticity is three times larger at 0.54, indicating large income effects.

The Marshallian and Hicksian elasticities show substantial heterogeneity. The 90-10 range of the Marshallian hours elasticity is 0.93 (from -0.14 to 0.79). For the Hicksian hours elasticity it is 0.78 (from 0.38 to 1.16). Differences in hours worked are an important source of variation in both the Hicksian and Marshallian elasticities. In Figure 3 we plot the full distribution of elasticities again by percentiles of the Marshallian distribution and according to hours worked. It is clear that those working the fewest hours show the largest proportional response to a wage increase.

In common with much of the labour supply literature, our estimates of the Marshallian elasticity are quite small (see Keane (2011) for a survey). They are comparable to those found for female labour supply in Blundell et al. (1998), who find values of the Marshallian elasticity ranging from 0.13 to 0.37 (depending on the age of the youngest child). Our estimates of the Hicksian elasticity are not too dissimilar from other estimates in the literature obtained a similar methodology to ours. For example, Blundell et al. (1998) obtain Hicksian elasticities ranging from just 0.14 to 0.44. Other results in the literature, however, report much larger estimates. MaCurdy (1983), for instance, estimates elasticities ranging from 0.74 to 1.43 (for men). The meta-study by Chetty et al. (2011) report an average elasticity (for men and women) of 0.33.

We investigated extensively the main reasons for different estimates of labour supply elasticities.

Our hypotheses ranged from the type of specification used,[26] to the type of variation in wages that is used to identify the elasticity (that is what type of instruments are used), to sample selection rules. As 2SLS or GMM approaches to estimate equilibrium conditions such as the MRS equation are sensitive to the normalization used, we also investigated whether the results we obtain depend on which variable is used as a dependent variable. It turns out that the normalization issue drives the result in a fundamental fashion, while results are robust to the other hypotheses considered. In particular, we find that IV or GMM estimates obtained using wages as the left hand side variable (as in MaCurdy) results in very large elasticities while putting leisure on the left hand side (similar to Blundell et al. (1998), who have hours worked as their dependent variable) yields very small elasticities. As noted above, we use a the Fuller estimator which is less sensitive to the normalisation of the estimating equation than these alternative methods. In Appendix B we report results from GMM estimation with different variables placed on the left hand side of the estimating equation.



Lines show the distributions of Marshallian, Hicksian and intensive Frisch elasticities smoothed using a local polynomial.

Figure 3: Intensive elasticities

6.2 Frisch hours elasticity

We use the estimates we obtain from the the Euler equation reported in section 5.2 to estimate the Frisch elasticities with respect to wages (at the intensive margin). Notice that these elasticities can

[26]That is whether one uses consumption to proxy for the marginal utility of wealth or other indicators.

Table 9: Elasticities at Percentiles of Marshallian distribution

	Wage			Interest rate
	Marshallian	Hicksian	Frisch	Frisch
Hours worked				
10th	−0.14 [−0.31,0.00]	0.38 [0.19,0.60]	0.80 [0.23,1.83]	0.78 [0.23,1.59]
25th	0.01 [−0.12,0.13]	0.44 [0.20,0.78]	0.80 [0.24,2.05]	0.76 [0.24,1.75]
50th	0.18 [0.05,0.37]	0.54 [0.24,1.07]	0.87 [0.24,2.35]	0.81 [0.24,1.92]
75th	0.39 [0.16,0.84]	0.69 [0.29,1.50]	1.00 [0.29,2.99]	0.93 [0.29,2.44]
90th	0.79 [0.38,1.69]	1.16 [0.55,2.32]	1.92 [0.58,4.67]	1.82 [0.58,3.86]
Leisure				
25th	−0.21 [−0.45,−0.09]	−0.38 [−0.83,−0.16]	−0.55 [−1.56,−0.16]	−0.51 [−1.27,−0.16]
50th	−0.11 [−0.23,−0.03]	−0.34 [−0.65,−0.16]	−0.55 [−1.44,−0.16]	−0.52 [−1.19,−0.16]
75th	−0.01 [−0.09,0.08]	−0.31 [−0.52,−0.14]	−0.55 [−1.37,−0.17]	−0.52 [−1.15,−0.17]
Consumption				
25th	0.82 [[0.68,1.08]	0.43 [0.18,0.87]	0.04 [−0.02,0.50]	−1.17 [−1.83,−0.54]
50th	1.05 [0.93,1.23]	0.52 [0.24,0.99]	0.05 [−0.02,0.58]	−1.19 [−1.84,−0.52]
75th	1.30 [1.14,1.46]	0.61 [0.30,1.07]	0.05 [−0.02,0.64]	−1.20 [−1.84,−0.50]

Elasticities are calculated as averages of the 10th, 25th, 50th, 75th and 90th percentiles of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

be obtained directly from the Euler equation using equations (24) and (25). These are shown in the third column of Table 9 and are plotted alongside Hicksian and Marshallian elasticities in Figure 3.

The Frisch elasticity for hours of work is larger than the Hicksian elasticity, as theory would predict. The elasticity also varies in the cross section rising from 0.8 at the 10th percentile of the Marshallian elasticity to 1.92 at the 90th percentile. The median value is 0.87. It is quite common to find large estimates of the Frisch hours elasticity among married women, and our findings are broadly in line with those of previous studies. In a recent paper, Blundell et al. (2016b) find a Frisch elasticity for married women of 0.96. Kimmel and Kniesner (1998) estimate a Frisch elasticity for this group of 0.67. Heckman and MaCurdy (1982) who also look at married women however obtain a much larger value for the Frisch of 2.35 (as reported in Keane (2011)).

Part of the heterogeneity we observe in the Frisch elasticities is due to differences across the life cycle and in demographics, but once again much of it is also due to differences in the level of hours of work. As with the Hicksian and Marshallian elasticities, Figure 3 shows that Frisch hours elasticities are largest for those working the fewest hours.

The elasticity of consumption with respect to anticipated wage changes is small but positive (owing to the fact consumption and leisure are substitutes). The Frisch elasticity of consumption with respect to the interest rate at the median level of consumption is -1.19.

We compare these results with those obtained when we impose additive separability for preferences over consumption and leisure, as well as when we use a standard CES utility specifications in Appendix E. This exercise highlights the importance of adopting a flexible utility specification. A standard CES specification, which is shown to be rejected by the estimation in section 5.1 leads to similar estimates of Marshallian hours elasticities, but much larger Hicksian and Frisch elasticities. The median Frisch hours elasticity estimated using the more restrictive standard CES specification is 1.33, which is roughly 50% larger than our baseline result. The corollary of this is that the Frisch with respect to the interest rate is much lower: imposing a standard CES forces consumption and leisure to have the same substitution parameters and so consumption is less elastic and leisure more elastic than in our baseline. In addition, the standard CES utility implies much greater non-separability between consumption and leisure: implying a Frisch wage elasticity of consumption of 0.4 compared to 0.05 under our more general utility specification.

6.3 The extensive margin and aggregate elasticities

The focus of the previous two subsections was on how responsive individuals' decisions over hours worked are to wage changes. This subsection reports on how responsive individuals' decisions about whether or not to participate are to wage changes. We approach this question by asking how much the percentage of women who work changes as the wage changes. We then calculate how total hours worked in the economy change as wages change, which is the result of both the extensive and intensive margin responses. This is what we call the "macro elasticity".

We explore the response to two different types of wage changes. First, in section 6.3.1, we focus on the response of labour supply to anticipated temporary changes in wages. This gives us the size of the intertemporal (Frisch) response at the extensive margin of labour supply. The response to anticipated temporary changes is the relevant one for temporary tax changes. We show how this response changes over the business cycle to show how responses to tax changes differ between booms and recessions. Second, in section 6.3.2 we calculate labour supply changes in response to changes in the entire life cycle wage profile. The size of this response is key in the discussion of the impact of different tax systems on labour supply and, as discussed in Chetty (2012), is conceptually similar to a Marshallian elasticity.

6.3.1 Response to anticipated temporary changes in wages

We first consider the various Frisch responses at different ages. Second we explore the heterogeneity in the responses of labor supply by looking at its variation across the household's wealth distribution and across demographic groups. Finally, we consider how the Frisch responses will vary over the business cycle.

Frisch responses are calculated by comparing labour supply in two economies that differ in a given quarter in that female wages are 10% higher for that quarter in one economy than in the other. This wage difference generates differences in participation rates, differences in hours worked for participants and, therefore, differences in total labour supply. In Table 10 we report the responses of labour supply at different ages. The 'extensive response' is calculated as the ratio of the percentage points variation in participation to the 10% wage variation (i.e., a 10% increase in wage produces 8.2 percentage points increase in participation in the age group 25-30) and it is reported in the first column. The second to forth columns report three different percentiles of the the hours intensive margin elasticity distribution at each age.[27] Finally, the last column reports the 'macro' elasticity: this is the change in the total number of hours worked, considering both intensive and extensive margins.

A first point to notice is the variation in the size of the extensive margin elasticity over the life cycle. As a consequence, the age composition of the population may have important implications for the aggregate response of labour supply to changes in wages. Early in life, the percentage point response is about 0.82, falling to 0.63 between 30 and 35 and to a minimum of 0.56 for the 40-45 group. The intensive margin median elasticity is pretty stable over the life cycle, however the 75th percentile elasticity increases substantially with age.

The macroeconomic elasticity that we compute is about 1.45 on average, but is larger at the start of

[27] The comparable value calculated direct from step 2 of the estimation process is 0.86. The similarity of estimates from step 2 and step 3 of the estimation provides validation for the more restrictive assumptions invoked in step 3.

Table 10: Frisch Responses

	Extensive Response (Percentage Pt)	Intensive Elasticity 25th	50th	75th	Macro Elasticity
25-30	0.82	0.71	0.85	1.09	1.85
30-35	0.63	0.67	0.83	1.10	1.48
35-40	0.64	0.65	0.82	1.13	1.45
40-45	0.56	0.65	0.86	1.20	1.35
45-50	0.59	0.66	0.88	1.24	1.39
50-55	0.59	0.68	0.91	1.28	1.45

The extensive response is the ratio of the percentage points variation in participation to the 10% wage variation.

the life cycle. The relative importance of the extensive and intensive margins to explaining the macro elasticity varies with age. Before age 30, the intensive margin response contributes approximately 46% of the response in the aggregate. However, by age 50-55, the contribution of the intensive response has increased to 63%. The contribution of the intensive margin is somewhat larger than Erosa et al. (2016), who find that the response through the intensive margin contributes about 38% to the aggregate response. This difference is not surprising since the Erosa et al. (2016) calculation is for men where we see less variability in hours worked. This does highlight the difficulty of aggregating behaviour to create a single labour supply elasticity.

Household Wealth. In Table 11, we explore the Frisch extensive and intensive margin responses across the household's wealth distribution. We calculate the percentiles of household's wealth at each age and classify households according to it in four different groups (those below percentile 25th, those between 25th and 50th percentiles, those between 50th and 75th percentiles and, finally, those above percentile 75th). We find a very clear pattern of a decreasing response of the extensive margin with increasing household's wealth. This is the case at all ages. There is also heterogeneity in the intensive margin elasticity by wealth, with the wealthy being less responsive, but the differences are more moderate than in the case of the extensive margin response. The message from these results is that the distribution of wealth is a key aspect to understand the response of aggregate labor supply to changes in wages.

Demographics. In Table 12, we explore the Frisch extensive margin response across women who differ in the age of youngest child living in the household. Mothers are more elastic than women without dependent children at each age group. And, except for the youngest mothers, the mothers

Table 11: Frisch Responses by Household Wealth

	Below $p25$	$p25 - p50$	$p50 - p75$	Above $p75$
	Extensive Margin Response			
25-30	1.62	1.02	0.46	0.16
30-35	1.08	0.91	0.44	0.09
35-40	1.26	0.66	0.39	0.24
40-45	1.12	0.59	0.37	0.15
45-50	1.01	0.78	0.40	0.16
50-55	1.22	0.69	0.28	0.15
	Intensive Margin Response			
25-30	1.18	0.94	0.81	0.71
30-35	1.19	1.01	0.77	0.66
35-40	1.17	1.00	0.76	0.65
40-45	1.21	0.97	0.79	0.67
45-50	1.20	1.07	0.83	0.68
50-55	1.22	1.17	0.89	0.69

of the youngest children are the most elastic. The finding that younger women are more elastic also holds for each of the three groups considered here.

Macroeconomic Conditions. We show now how the response of the extensive margin may change across different scenarios in terms of macroeconomic conditions. Differences in the economic environment will lead to differences in the estimated elasticity for the same underlying preference parameters, as also discussed by Keane and Rogerson (2012). This issue is likely to be relevant particularly for the extensive margin, which is driven by non-convexities in the dynamic problem, such as fixed costs of going to work. If these non-convexities are important, it is likely that a certain sequence of aggregate shocks will tend to bunch (or further disperse) households around the kinks that determine the extensive margin response. As a consequence, different distributions of the state variables will trigger different responses in the aggregate. In particular, whether an economy is in a recession or not may well affect how much individuals are willing to respond to wage growth. This is what makes it particularly difficult to think of a single labour supply elasticity.

In Table 13, we analyse the extensive margin labour supply responses of women over the life cycle to deterministic changes in wages at different points of the business cycle to highlight how the state

Table 12: Frisch Extensive Margin Response, Demographic groups

	Mothers 0-2	Mothers 3-17	Without Dependent Children
25-30	0.92	0.96	0.65
30-35	0.81	0.64	0.55
35-40	0.73	0.66	0.53

Table 13: Frisch Extensive Margin Response, Business Cycle

	Baseline	Recession		
		Average	Q1	Q4
25-30	0.82	0.86	0.85	0.89
30-35	0.63	0.72	0.68	0.78
35-40	0.64	0.77	0.72	0.84
40-45	0.56	0.67	0.64	0.71
45-50	0.59	0.69	0.66	0.71
50-55	0.59	0.69	0.69	0.74

of the economy affects Frisch labour supply responses.[28] We report the labour supply response in the first and fourth quarters of the recession, as well as the average response in a recession. The key finding is that responses are higher in recessions than in the baseline, and further, responses increase with the duration of the recession. From the results in Table 11, the decrease in wealth that households suffer over a recession could be behind the increasing responsiveness of the extensive margin to anticipated changes in wages.

These results show responsiveness while in a recession. The effects of recessions on responsiveness may persist beyond the end of the recession, especially if wages are permanently lower or wealth is lower. Both lower wages and lower wealth lead to higher elasticities, and so households who have been hit by recessions earlier in their life are more responsive throughout the remainder of their lives.[29]

[28] In the simulation used to derive these numbers, we define a recession as a situation in which all men and women receive an unexpected negative earnings shock for four consecutive quarters. These wage changes are to the permanent wage and will affect the marginal utility of wealth as well as changing intertemporal incentives.

[29] We show this by using our simulations to compare women hit by a recession at age 25 with those not hit by recession. Differences persist throughout their lifetimes.

6.3.2 Long-run responses to changes in wage profiles

The response to an unanticipated change to the entire wage profile is the response to an unanticipated permanent tax change that is unfunded. It is a similar concept to the Marshallian elasticity but in a life cycle context. This is the type of response of interest in the discussion of differences in labour supply across countries with different tax systems or in steady state comparisons. As discussed by Keane (2011), in considering the implications of different tax systems, the response depends on what tax revenue is used for and how these uses enter the utility function. We consider the extensive, intensive margin and the macro response to an increase in the entire wage profile of 10% for both husband and wife. Simulated labour supply responses to unanticipated changes to the wage profile are muted. The median intensive margin response is 0.13. As noted by Meghir and Phillips (2008) this elasticity concept is different from the Marshallian elasticity we measure using MRS. In the current experiment, the intensive margin response incorporates the effect of changes in savings that takes place as a result of the changes in wages, in addition to the response if savings are kept constant within the period. The intensive margin response is slightly smaller than the 0.18 Marshallian we estimated using the MRS. The extensive margin response is also limited, with a value of 0.23, which is a third of the Frisch extensive response in Table 10. Of course, if we consider a scenario in which only women's wages are increased by 10% but husbands' earnings are kept constant, the response of labour supply is larger because of the absence of an income effect through the husband's earnings. In this case, the extensive margin response rises to 0.51.

7 Returns to experience

An important maintained assumption to this point has been the absence of any returns to experience. Imai and Keane (2004) argue that assuming wages are exogenous may introduce a downward bias in the estimates of the willingness to substitute intertemporally. Indeed, they present estimates of such a parameter as high as 3.8 in a model that accounts for returns to labor market experience.[30]

In this section, we consider an alternative framework in which returns to experience accrue to individuals who are participating, but these returns are not affected by the number of hours worked conditional on participation. In particular, we assume that human capital accumulates according to the following process:

$$\ln e^f_{h,t} = \ln e^f_{h,t-1} + \nu I\left(P_{h,t-1} = 1\right) - \delta I\left(P_{h,t-1} = 0\right)$$

We make the assumption that returns to experience operate only through the participation margin because in this case, the estimates of the MRS and Euler equations remain valid. However, we need to change the solution for the discrete choices.

[30] However, as discussed in Wallenius (2011), Imai and Keane (2004) base their identification on the early periods of the life cycle. Their model does a less good job of accounting for the life cycle profile at later ages using these estimates.

We begin by recalibrating the parameter values that were chosen in the baseline economy to fit some of the features of participation: the fixed cost of working, $\bar{F}$, child care price, p, the offered wage gender gap and ψ_0. In addition to these parameters, we also need to calibrate the parameter that characterises human capital accumulation function and its depreciation rate.[31] In order to identify all these parameters we target the female participation rate, the participation rate of mothers, the average hours worked, the observed wage gender gap, the observed wage growth at early ages, and the observed depreciation of wages during non-participation. Note that the value of the statistics on wages are shaped by selection so we need to identify the underlying parameters by solving the model. We report the implied parameters in Table 14. As expected, in the context of returns to experience, a much larger fixed cost of working and childcare cost are required in order to match participation statistics.

Appendix F provides additional statistics analogous to the baseline model. In particular, analogously to Figure 2, Figure 4 shows life cycle profiles in the simulations and in the data; and Table 26 reports additional statistics on the distribution of hours and of wages.

7.1 Response to anticipated temporary changes in wages

In Table 15, we report the labour supply responses in the economy with returns to experience. The key finding is that, in contrast to the economy without returns to experience, the extensive margin response is essentially zero and, as a result, the macro elasticity is about half of the one in the baseline economy (that we report in the last column for the sake of comparison). In this economy, there is a strong incentive to participate to obtain the return to experience. The larger fixed cost of participating that is estimated in this economy alongside the strong incentive to participate implies that changes in the current wage make little difference to the incentive to participate. As expected, the size of the intensive margin response is similar to the one in the economy without returns to experience. As we argued above, the extensive margin response is not governed by the estimated preference parameters from the MRS, and the response is very different in the two setups that we consider. Our results here are similar to Imai and Keane (2004) who argue that the response of labor supply to changes in wages may be mitigated in the context of endogenous wages.

It may well be that the small response of the extensive margin labour supply that we find is related to the simple model of return to experience we have considered. Whether returns to experience operate in a more subtle manner through intensive margins and the number of hours is a question we leave for future research. If that is the case, we would need to change substantially the estimation methods we used in the first part of the paper.

One possibility, of course, is that returns to tenure are important for some occupations and/or skill

[31] Note this is only one parameter in contrast to the two parameters ι_1^f and ι_2^f for the exogenous wage growth that were used in the baseline economy.

Table 14: Baseline economy: Calibrated Parameters and Targets

Parameter Name		Values	
		Ret to Exp	Baseline
Constant term weight of leisure	ψ_0	4.13	4.20
Childcare Cost	p	5820	967
Fixed Cost of Working	$\bar{F}$	315	468
Offered Wage Gender Gap at age 22	y_0^f/y_0^m	0.79	0.74
Standard Deviation of Permanent Shock (Women)	σ_{ξ^f}	0.063	0.063
Standard Deviation of Initial Wage (Women)	$\sigma_{\xi^f,0}$	0.50	0.50
Exogenous growth in offered wage	ι_1^f	-	0.052
Exogenous growth in offered wage	ι_2^f	-	-0.0006
Female Human Capital Tech	ν	0.003	-
Discount Factor (annualized)	β	0.99	0.99

Targets	Data	Ret to Exp	Baseline
Weekly hours worked	37.3	37.0	37.3
Participation Rate	0.677	0.690	0.678
Participation Rate of Mothers	0.538	0.544	0.546
Observed Wage Gender Gap	0.720	0.720	0.727
Observed Variance Wage Growth (Women)	0.005	0.005	0.005
Observed Initial Variance of Wages (Women)	0.15	0.15	0.15
Wage Growth (if younger than 40)	0.012	0.013	0.010
Wage Growth (if older than 40)	0.001	0.013	0.004
Median wealth to income ratio	1.84	1.82	1.80
Observed Depreciation Rate	-0.050	-0.040	0.02

Statistics for women born in the 1950s and aged 25 to 55. Wage growth is annual.

Table 15: Returns to experience: Frisch Changes

	Extensive Response	Intensive Elasticity 25th	50th	75th	Macro Elasticity	Baseline
25-30	0.02	0.65	0.81	1.15	0.91	1.85
30-35	0.04	0.63	0.79	1.17	0.91	1.48
35-40	0.03	0.63	0.78	1.17	0.90	1.45
40-45	0.03	0.61	0.79	1.19	0.89	1.35
45-50	0.04	0.60	0.77	1.19	0.88	1.39
50-55	0.07	0.58	0.75	1.09	0.86	1.45

The extensive response is the ratio of the percentage points variation in participation to the 10% wage variation.

levels and not for others. In such a case, it would be necessary to introduce an additional dimension of heterogeneity that would make the aggregation issues we have repeatedly stressed even more salient.[32]

7.2 Long-run responses to changes in wage profiles

Finally, in Table 16 we report the extensive, intensive margin and the macro responses to an increase in the entire wage profile of 10% for both husband and wife. In this case the response both at the extensive and the intensive margin is very similar in the economy with and without returns to experience.

Table 16: Labour supply changes, Marshallian

	Extensive Response	Intensive Elasticity 25th	50th	75th	Macro Elasticity
Ret to experience	0.22	0.07	0.10	0.17	0.33
Baseline	0.23	0.10	0.13	0.19	0.35

The extensive response is the ratio of the percentage points variation in participation to the 10% wage variation.

[32]Alternatively it could be that returns to experience depend on hours worked. Blundell et al. (2016a) show that these returns are close to zero for part-time work.

8 Conclusion

This paper shows that in understanding labour supply behaviour and in calculating aggregate labour supply elasticities, heterogeneity across individuals is very important. We also stress that aggregate elasticities can vary substantially over the business cycle and with the duration of recessions. To make these points precisely and show their quantitative importance, we estimate a life cycle model of work and saving choices and characterise the response of female labour supply to different types of wage changes. In the process of estimating such a model, we use a flexible specification of preferences that allows us to test some of the assumptions commonly used in the macro literature on labour supply. Empirically, we use as a robust an approach as possible.

We find substantial heterogeneity in labour supply responses, and this heterogeneity is prevalent at both the intensive and extensive margins. The median Marshallian elasticity is 0.18, but has an interquartlie variation of 0.01 to 0.39 and *90-10* range of -0.14 to 0.79. The corresponding Hicksian elasticity is 0.54, with interquartile variation of 0.44 to 0.69 and *90-10* range of 0.38 to 1.16; and the corresponding Frisch wage elasticity is 0.87, with an interquartile range from 0.8 to 1.0 and *90-10* range of 0.8 to 1.92. In terms of heterogeneity in the intensive margin due to observable characteristics, the Marshallian, Hicksian and Frisch elasticities are greatest for those working the least number of hours, and those with the least wealth. For the extensive margin, the response to anticipated wage growth is large for women under 30, and for those with young children, and can explain 54% of their labour supply response. This sizable contribution of the extensive margin declines with age. We find some evidence of nonseparability between consumption and leisure, but assuming there is separability does not substantially change the estimates of the distribution of the Frisch elasticity.

Our preference parameter estimates reject the restrictions required for balanced growth, which are widely used in the macro literature. The curvature on consumption in utility is less than log, and the curvature on hours worked is much greater than the curvature on consumption. This implies individuals are less willing to substitute hours of work over time than they are willing to substitute consumption. Further, the measured heterogeneity means it is not sensible to talk about a single elasticity measuring how labour supply in the economy responds to wage changes. Instead, one can aggregate explicitly from individual behaviour to the aggregate to understand how economy wide hours of work change given the demographic and age structure of the economy, the wealth distribution and the state of the business cycle.

We use the estimates of our model and the micro data we use to aggregate explicitly individual behaviour and construct a consistent estimate of a 'macro' labour supply elasticity, which we define as the percentage change in aggregate hours corresponding to a given change in net wages. Our model allows to decompose such an aggregate elasticity in the parts due to the extensive and the intensive margins.

Our results on the importance of the extensive margin in explaining macro elasticities can be compared to others in the literature, especially Erosa et al. (2016) and Guner et al. (2012). Our estimates put a greater importance on intensive margin changes in hours worked per week than those papers, but we do find that a substantial fraction of the changes in total hours is due to changes in participation, ranging from 54% to 37%. Erosa et al. (2016) find that the extensive margin is the dominant labour supply response, explaining 62% of the aggregate response. Their model has a similar life cycle structure to ours, but is focused on male labour supply and the conclusion on the importance of the extensive margin is for men where hours of work are less variable. Guner et al. (2012) analyse the importance of the female extensive margin for the aggregate response of labor supply to changes in taxes in a model with heterogeneous married and single households, and with a female extensive margin as well as a male and female intensive margin. As with Erosa et al. (2016), they find that the female extensive margin is a key contributor to the aggregate response to tax reform. The key difference from our framework is their assumption that there is no uncertainty in wages and this assumption tends to lead to greater labour supply responses, as shown in Low (2005).

One key point that emerges from our exercise is that aggregate responses of labour supply to changes in wages (both at the intensive and the extensive margin) is not constant: it changes with the structure of the population as well as with the state of the economy. This finding is similar to Keane and Rogerson (2012), who argue that there is no contradiction between macro and micro elasticities of labour supply and that they are simply measuring different concepts. Our conclusion is however stronger: the macro elasticity is not a structural parameter, it is simply the result of highly nonlinear aggregation which depends on demographic structure as well as the distribution of wealth and the particular point in the business cycle.

References

Aaronson, D. and E. French (2004): "The Effect of Part-Time Work on Wages: Evidence from the Social Security Rules," *Journal of Labor Economics*, 22, 329–352.

Abowd, J. M. and D. Card (1989): "On The Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411–445.

Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 17, 36–49.

Attanasio, O. P. (1999): *Handbook of Macroeconomics*, Elsevier, vol. 1, chap. 11, 741–812.

Attanasio, O. P. and H. Low (2004): "Estimating Euler Equations," *Review of Economic Dynamics*, 7, 406–435.

Attanasio, O. P., H. Low, and V. Sánchez-Marcos (2008): "Explaining Changes in Female Labor Supply in a Life-cycle Model," *American Economic Review*, 98, 1517–1552.

Attanasio, O. P. and G. Weber (1995): "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from The Consumer Expenditure Survey," *Journal of Political Economy*, 103, 1121–1157.

——— (2010): "Consumption and Saving: Models of Intertemporal Allocation and their Implications for Public Policy," *Journal of Economic Literature*, 48, 693–751.

Blundell, R., A. Bozio, and G. Laroque (2011): "Labor Supply and the Extensive Margin," *American Economic Review*, 101, 482–486.

Blundell, R., M. Costa, C. Meghir, and J. Shaw (2016a): "Female Labor Supply, Human Capital and Welfare Reform," *Econometrica*, 84, 17051753.

Blundell, R., A. Duncan, and C. Meghir (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66, 827–861.

Blundell, R. and T. MaCurdy (1999): *Handbook of Labor Economics*, Elsevier, vol. 3, chap. Labor supply: A review of alternative approaches, 1559–1695.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016b): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106, 387–435.

Blundell, R. and I. Walker (1986): "A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data," *Review of Economic Studies*, 53, 539–558.

Browning, M., A. Deaton, and M. Irish (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53, 503–544.

Browning, M., L. P. Hansen, and J. J. Heckman (1999): "Micro data and general equilibrium models," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, Elsevier, vol. 1 of *Handbook of Macroeconomics*, chap. 8, 543–633.

Chang, Y. and S.-B. Kim (2006): "From individual to aggregate labor supply : a quantitative analysis based on a heterogeneous agent macroeconomy," *International Economic Review*, 47, 1–27.

Chetty, R. (2012): "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review*, 101, 471–475.

Chetty, R., A. Guren, D. S. Manoli, and A. Weber (2013): "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, 27, 1–56.

Davidson, R. and J. G. MacKinnon (2004): *Econometric Theory and Methods*, Oxford University Press.

Deaton, A. (1985): "Panel Data from Time Series of Cross-Sections," *Journal of Econometrics*, 30, 189–194.

Erosa, A., L. Fuster, and G. Kambourov (2016): "Towards a micro-founded theory of aggregate labor supply," *Review of Economic Studies*, 83, 1001–1039.

Feenberg, D. and E. Coutts (1993): "An introduction to the TAXSIM model," *Journal of Policy Analysis and Management*, 12, 189–194.

Fuller, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–953.

Goldin, C. and L. F. Katz (2007): "Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing," *Brookings Papers on Economic Activity*, 2, 135–167.

Gorman, W. M. (1959): "Separable utility and aggregation," *Econometrica*, 27, 469481.

Guner, N., R. Kaygusuz, and G. Ventura (2012): "Taxation and Household Labour Supply," *Review of Economic Studies*, 79, 11131149.

Hahn, J. and J. Hausman (2002): “A New Specification Test For The Validity Of Instrumental Variables,” *Econometrica*, 70, 163–189.

——— (2003): “Weak Instruments: Diagnosis and Cures in Empirical Econometrics,” *American Economic Review*, 93, 118–125.

Hahn, J., J. Hausman, and G. Kuersteiner (2004): “Estimation with Weak Instruments: Accuracy of Higher-order Bias and MSE Approximations,” *Econometrics Journal*, 7, 272–306.

Hansen, L. P., J. Heaton, and A. Yaron (1996): “Finite-Sample Properties of Some Alternative GMM Estimators,” *Journal of Business & Economic Statistics*, 14, 262–280.

Hausman, J., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): “Instrumental Variable Estimation with Heteroskedasticity and Many Instruments,” *Quantitative Economics*, 3, 211–255.

Hayashi, F. (1987): “Tests for Liquidity Constraints: A Critical Survey,” in *Advances in Econometrics, Fifth World Congress*, ed. by T. Bewley, Cambridge University Press, vol. 2, 91–120.

Heckman, J. J. and T. E. Macurdy (1980): “A Life Cycle Model of Female Labour Supply,” *The Review of Economic Studies*, 47, 47–74.

Heckman, J. J. and T. E. MaCurdy (1982): “Corrigendum on a Life Cycle Model of Female Labour Supply,” *Review of Economic Studies*, 49, 659–660.

Hirsch, B. T. (2005): “Why Do Part-Time Workers Earn Less? The Role of Worker and Job Skills,” *ILR Review*, 58, 525–551.

Huggett, M., G. Ventura, and A. Yaron (2011): “Sources of Lifetime Inequality,” *American Economic Review*, 101, 2923–2954.

Hyslop, D. R. (2001): “Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings,” *American Economic Review*, 91, 755–777.

Imai, S. and M. P. Keane (2004): “Intertemporal Labor Supply and Human Capital Accumulation,” *International Economic Review*, 45, 601–641.

Keane, M. P. (2011): “Labor Supply and Taxes: A Survey,” *Journal of Economic Literature*, 49, 961–1075.

Keane, M. P. and R. Rogerson (2012): “Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom,” *Journal of Economic Literature*, 50, 464–476.

——— (2015): "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective," *Annual Review of Economics*, 7, 89–117.

Kimmel, J. and T. J. Kniesner (1998): "New Evidence on Labor Supply: Employment versus Hours Elasticities by Sex and Marital Status," *Journal of Monetary Economics*, 42, 289–301.

Ljungqvist, L. and T. J. Sargent (2011): "A Labor Supply Elasticity Accord?" *American Economic Review*, 101, 487–491.

Low, H. (2005): "Self-insurance in a Life-cycle Model of Labour Supply and Savings," *Review of Economic Dynamics*, 8, 945–975.

MaCurdy, T. (1981): "An Empirical Model of Labour Supply in a Life Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

——— (1983): "A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty," *International Economic Review*, 24, 265–289.

Mariano, R. and T. Sawa (1972): "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, 67, 159–163.

Meghir, C. and D. Phillips (2008): "Labour Supply and Taxes," *IZA WP*, 3405.

Miller, R. A. and H. Sieg (1997): "A Microeconometric Comparison of Household Behavior between Countries," *Journal of Business & Economic Statistics*, 15, 237–253.

Nelson, C. R. and R. Startz (1990a): "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One," *Journal of Business*, 63, S125–S140.

——— (1990b): "Some Further Results on the Exact Small Sample Properties of the Instumental Variables Estimator," *Econometrica*, 58, 967–976.

Rendall, M., E. Aracil, C. Bagavos, C. Couet, A. DeRose, P. DiGiulio, T. Lappegard, I. Robert-Bobée, M. Rønsen, S. Smallwood, and G. Verropoulou (2010): "Increasingly Heterogeneous Ages at First Birth by Education in Southern European and Anglo-American Family-Policy Regimes: A Seven-country Comparison by Birth Cohort," *Population Studies*, 64, 209–227.

Rogerson, R. and J. Wallenius (2009): "Micro and Macro Elasticities in a Life Cycle Model with Taxes," *Journal of Economic Theory*, 144, 2277–2292.

Sawa, T. (1972): "Finite-Sample Properties of the k-Class Estimators," *Econometrica*, 40, 653–680.

Staiger, D. and J. H. Stock (1997): "IV Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, chap. 6, 80–108.

Wallenius, J. (2011): "Human Capital Accumulation and the Intertemporal Elasticity of Substitution of Labor: How Large is the Bias?" *Review of Economic Dynamics*, 14, 577–591.

Ziliak, J. P. and T. J. Kniesner (1999): "Estimating Life-cycle Labor Supply Tax Effects," *Journal of Political Economy*, 107, 326–359.

Online Appendix A: Elasticity derivations

In this section we provide the formulae for the first and second derivatives that are used to calculate the different elasticities. We define $D = \exp(\pi z + \xi P + \zeta)$ (omitting subscripts for convenience). Then it is easy to show that:

$$u_c\left(c,l\right) = DM^{-\gamma}c^{-\phi} \tag{31}$$

$$u_l\left(c,l\right) = D\alpha M^{-\gamma}l^{-\theta} \tag{32}$$

$$u_{cl}\left(c,l\right) = (-\gamma)DM^{-\gamma-1}\alpha c^{-\phi}l^{-\theta} \tag{33}$$

$$u_{ll}\left(c,l\right) = (-\gamma)\frac{u_l\left(c,l\right)}{\alpha M}l^{-\theta} - u_l\left(c,l\right)\theta l^{-1} \tag{34}$$

$$u_{cc}\left(c,l\right) = (-\gamma)\frac{u_c\left(c,l\right)}{M}c^{-\phi} - u_c\left(c,l\right)\phi c^{-1} \tag{35}$$

Finally, note that:

$$u_{cl}\left(c,l\right) = (-\gamma)u_c\left(c,l\right)l^{-\theta}\frac{\alpha}{M} = (-\gamma)u_l\left(c,l\right)c^{-\phi}\frac{1}{M} \tag{36}$$

These expressions can be used to calculate the Frisch elasticities in the paper. The formula for the wage Frisch for intensive margin choices can be derived as follows:

$$\begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix} \begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \frac{1}{u_{cc}u_{ll} - u_{cl}^2} \begin{bmatrix} u_{ll} & -u_{cl} \\ -u_{cl} & u_{cc} \end{bmatrix} \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\varepsilon_c^F = \frac{w}{c}\frac{\partial c}{\partial w} = -\frac{u_c u_{cl}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{c}$$

$$\varepsilon_l^F = \frac{w}{l}\frac{\partial l}{\partial w} = \frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{l}$$

$$\varepsilon_h^F = \frac{w}{h}\frac{\partial h}{\partial l}\frac{\partial l}{\partial w} = -\frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{h} = -\varepsilon_l\frac{l}{h}$$

The formula for the interest-rate Frisch can similarly be derived as follows:

$$\begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix} \begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix}^{-1} \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \frac{1}{u_{cc}u_{ll} - u_{cl}^2} \begin{bmatrix} u_{ll} & -u_{cl} \\ -u_{cl} & u_{cc} \end{bmatrix} \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\varepsilon_c^{F_R} = \frac{(1+R_{t+1})}{c} \frac{\partial c}{\partial(1+R_{t+1})} = \frac{u_c u_{ll} - u_l u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_c u_{ll} - u_l u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)}$$

$$\varepsilon_l^{F_R} = \frac{(1+R_{t+1})}{l} \frac{\partial l}{\partial(1+R_{t+1})} = \frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_l u_{cc} - u_c u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)}$$

$$\varepsilon_h^{F_R} = \frac{(1+R_{t+1})}{h} \frac{\partial h}{\partial l} \frac{\partial l}{\partial(1+R_{t+1})} = -\frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{h} = -\varepsilon_l^{F_R} \frac{l}{h}$$

Online Appendix B: Alternative methods of estimating the MRS

In this appendix we discuss results from alternative MRS specifications. For comparison with later results, we present a fuller set of parameter estimates from our baseline MRS specification in Table 17.

Estimation method and normalisation

We start by considering the issue of how the MRS is normalised. Recall that our MRS relationship is

$$\ln w_{h,t} = \psi_0 + \psi z_{h,t} - \theta \ln l_{h,t} + \phi \ln c_{h,t} + \upsilon_{h,t} \tag{37}$$

As Keane (2011) notes, this is not a labour supply equation but an equilibrium condition in which wages, leisure and consumption are all endogenous. All three variables are potentially correlated with the error term $\upsilon_{h,t}$ and so there is no natural choice of the dependent variable.

Despite this, we find that, when conventional methods are used, results can be highly sensitive to whether wages, leisure or consumption are placed on the left hand side of the MRS equation. Table 18 shows results from estimating ϕ and θ using GMM under the three different possible normalisations. It is clear that our parameter estimates and elasticities vary a great deal depending on which normalisation is used. When wages are selected as the left-hand side variable, elasticites are relatively large.

Table 17: Baseline MRS estimates

Parameter	Estimate	(Standard Error)	[95% Confidence Interval]
θ	1.75**	(1.230)	[0.34,5.12]
ϕ	0.76***	(0.103)	[0.55,0.95]
Ψ			
Age	0.05*	(0.02)	[0.01,0.09]
Age^2	-0.0005	(0.0007)	[-0.003,0.001]
Age^3/1000	-0.01	(0.01)	[-0.03,0.03]
Age^4/10000	0.002	0.0007	[-0.0005,0.003]
North East	0.01	(0.03)	[-0.06,0.14]
Mid West	-0.05	(0.01)	[-0.08,0.00]
South	-0.11	(0.02)	[-0.22,-0.05]
White	-0.04	(0.03)	[-0.13,0.08]
No. elderly HH members	0.02	(0.02)	[-0.05,0.06]
ln($famsize$)	-0.32***	(0.037)	[-0.38,-0.23]
Has kids	0.07***	(0.021)	[0.04, 0.10]
No. of kids 0-2	0.15***	(0.030)	[0.10, 0.22]
No. of kids 3-15	0.06***	(0.017)	[0.04, 0.10]
No. of kids 16-17	-0.02**	(0.011)	[-0.05,0.00]
Constant (Ψ_0)	4.70	(4.94)	[-8.74,28.03]
Heckman selection terms			
e_1	0.07	(0.167)	[-0.18, 0.48]
e_2	0.05	(0.172)	[-0.21, 0.51]
e_3	0.01	(0.052)	[-0.08, 0.13]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Additional controls for season and year dummies and cohort-education interactions. Confidence intervals are bootstrapped with 1000 replications allowing for clustering at the individual level.

When leisure is the dependent variable, they are much smaller. Very similar considerations apply to the estimation of our Euler equation.

Differences of this kind can emerge in IV estimation in 2SLS and GMM estimation when the instruments chosen are relatively weak. Indeed, Hahn and Hausman (2002, 2003) propose using the differences in parameters implied by 2SLS estimates run under different normalisations as a test of instruments' strength.

Various papers have discussed possible remedies for cases when strong instruments are not available (Hahn and Hausman, 2003; Hausman et al., 2012). One possible solution is the use estimators such as Limited Information Maximum Likelihood (LIML) rather than 2SLS, which is known to have poor bias properties in such circumstances (Staiger and Stock, 1997; Nelson and Startz, 1990a,b). Using the notation from Davidson and MacKinnon (2004), for the case where

Table 18: MRS Estimates using GMM

	Dependent variable		
	Wages	Leisure	Consumption
Parameters			
θ	0.46^{**}	-13.8	0.13
	[−0.03,0.61]	[−86.53,154.63]	[−0.55,0.54]
ϕ	0.61^{***}	0.17	1.38^{***}
	[0.48,0.65]	[−4.09,0.12]	[1.24,1.73]
Wage elasticities at median			
Marshallian	0.55	0.09	0.17
	[0.52,1.13]	[0.00,0.13]	[−0.40,−0.07]
Hicksian	1.19	0.11	0.77
	[1.10,2.15]	[−0.10,0.13]	[0.60,1.02]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 4. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

$$
\begin{aligned}
y &= Z\beta_1 + Y\beta_2 + u = X\beta + u \\
Y &= \Pi W + v
\end{aligned}
$$

where $W = [ZW_1]$ (with W_1 being a matrix of instruments), and matrices X and W are $n \times k$ and $n \times l$ respectively (with $l \geq k$). The LIML estimator can be written as

$$\hat{\beta}^{\text{LIML}} = (X'(I - k_{\text{LIML}} M_W)X)^{-1} X'(I - k_{\text{LIML}} M_W)y \tag{38}$$

where $M_W = I - W(W'W)W'$, and

$$k_{\text{LIML}} = \frac{(y - Y\beta_2)' M_Z (y - Y\beta 2)}{(y - Y\beta_2)' M_W (y - Y\beta 2)} \tag{39}$$

However, while LIML is often found to have better bias properties than 2SLS, it has long been recognised that LIML does not have moments in finite samples (Mariano and Sawa, 1972; Sawa, 1972), and simulation exercises have shown that this can add considerable volatility to empirical estimates (Hahn et al., 2004). As a result Hahn et al. (2004) recommend the use of either jack-knifed 2SLS or the modification of LIML proposed by Fuller (1977). For this latter estimator, we replace k_{LIML} in equation (38) with

$$k_{\text{Fuller}} = k_{\text{LIML}} - \frac{\lambda}{(n-k)} \tag{40}$$

where λ here is a parameter chosen by the researcher, to obtain a value for $\hat{\beta}^{\text{Fuller}}$. The resulting estimator is guaranteed to have moments in finite samples Fuller (1977). We choose a value of one for this as suggested by Davidson and MacKinnon (2004) as it yields estimates that are approximately unbiased.

As well as its superior bias properties, the Fuller estimator has the advantage that is much less sensitive than GMM or 2SLS to the choice of the dependent variable, as Table 19 shows. Both the elasticity and parameter estimates obtained using alternative normalisations of the Fuller estimator are very similar to our baseline results.

Table 19: MRS Estimates with Different Dependent Variables

	Dependent variable	
	Leisure	Consumption
Parameters		
θ	1.84	1.75
	[−15.81,19.48]	[−6.25,9.74]
ϕ	0.76	0.77
	[−0.23,1.75]	[−0.22,1.76]
Wage elasticities at median		
Marshallian	0.17	0.18
	[−0.04,0.38]	[−0.02,0.37]
Hicksian	0.53	0.54
	[−0.06,1.11]	[0.02,1.07]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 4. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

Alternative instruments

In Table 20 we show results using alternative choices of instruments. We show results using GMM (with wages as the dependent variable) and the Fuller estimator described above. The first two columns show results with two alternative sets of instruments: using region-education-year interactions or using a *full* set of cohort-education-year interactions as used in Blundell et al. (1998). The former approach exploits variation in the wages and hours of workers with different education levels across regions and over time rather than variation across cohorts. The latter is similar to the approach we adopt for our main results but interacts cohort-eduction dummies full set of year effects rather than a polynomial in time trends.

The choice of instruments does have some impact on the results. Using variation in regions and education groups we obtain somewhat smaller values of the Marshallian and Hicksian elasticities. As we noted in section 4.1 however, estimates from such an approach might be contaminated by changes in the composition of education groups over time. We thus prefer the approach that we adopt in our main specification. The estimates we obtain from fully adopting the Blundell et al. (1998) approach are very similar to our main results, though somewhat less precise.

The sensitivity or our results to the choice of instruments is on the whole quite small when we compare it to the differences that can arise from the choice of estimation method. Just as we find for our main set of results, the hours elasticities estimated using the GMM estimator with wages as the dependent variable are substantially larger than those using the Fuller estimator for both of our alternative instrument sets.

Table 20: MRS Estimates using alternative instruments

	Region-Education-Year dummies		BDM (1998)	
	Fuller	GMM	Fuller	GMM
Parameters				
θ	5.09***	0.63***	1.93	0.08
	[1.19,9.00]	[0.44,0.82]	[−5.75,9.61]	[−0.12,0.27]
ϕ	0.73***	0.25***	0.76***	0.52***
	[0.27,1.19]	[0.21,0.29]	[0.48,1.04]	[0.46,0.52]
Wage elasticities at median				
Marshallian	0.09	1.22	0.17	1.08
	[−0.01,0.19]	[0.65,1.79]	[−0.43,0.77]	[0.53,1.62]
Hicksian	0.25	1.57	0.51	2.92
	[0.14,0.36]	[0.90,2.25]	[−0.86,1.88]	[1.06,4.79]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. BDM (1998) instruments are a full set of cohort-education-year dummies. Controls as in Table 4 except that in the region-education-year specification we replace cohort-education interactions with education dummies. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

Alternative samples

Table 21 shows how our MRS results are affected by alternative sample selection choices. Column (1) presents results when we exclude those individuals who report working exactly 40 hours a week. The justification of this experiment is that these individuals may be affected by some kind of friction that does not allow them to adjust their hours worked as desired. Such frictions would mean that the MRS condition that we exploit to recover ϕ and θ need not hold, and so it may be more reasonable to exclude these observations. When we do so, we obtain greater estimates of our Marshallian and Hicksian hours elasticities (at 0.45 and 0.72 respectively). These values are however somewhat imprecisely estimated and the confidence bands that surround them include our baseline estimates.

In Column (2) we show results when we exclude individuals working less than 20 hours per week (with an appropriate adjustment to our selection correction). We consider results from this specification because there may be certain frictions that prevent individuals working fewer hours than this, which would again lead to potential violations of the MRS condition. Excluding these observations delivers somewhat lower elasticity estimates, but again the estimates are imprecise.

Table 21: MRS Estimates using alternative samples

	Exc. 40 hours	Exc. <20 hours	Born 1925-1965
	(1)	(2)	(3)
Parameters			
θ	1.52	2.81	2.08**
	[−2.13,5.18]	[−2.69,8.32]	[0.05,4.10]
ϕ	0.42	0.76***	0.56***
	[−0.05,0.90]	[0.42,1.09]	[0.36,0.76]
Wage elasticities at median			
Marshallian	0.45	0.13	0.27
	[−0.39,1.29]	[−0.06,0.32]	[−0.04,0.58]
Hicksian	0.72	0.39	0.53
	[−0.72,2.17]	[−0.26,1.05]	[0.04,1.03]
N	26,060	47,743	39,057

*p<0.10, ** p<0.05, *** p<0.01. Specification (1) excludes individuals who work exactly 40 hours. Specification (2) excludes those working less than 20 hours (part-time workers). Specification (3) only includes individuals from cohorts with the most similar labour supply choices over the life cycle. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

Finally, in Column (3) we consider only those ten-year birth cohorts with the most similar labour-supply behaviour over the life cycle. In particular we exclude those born before 1925 as they tend to work fewer hours at older ages than other cohorts, and those born after 1975, as less-educated individuals born after this date tend to have lower employment rates than other earlier cohorts at the same ages. Using this sample, we obtain a Marshallian elasticity of 0.27 and a Hicksian 0.53. While the Marshallian elasticity estimated from this sample is slightly higher than our baseline estimates, the Hicksian elasticity is essentially unchanged.

Alternative definitions of hours

In Table 22 we consider how elasticity estimates are affected when we use an alternative measure of hours of leisure. The measure we use here is

$$\text{leisure} = \frac{5200 - \text{hours per week} \times \text{weeks worked per year}}{52} \tag{41}$$

This measure accounts for the observed variation in weeks worked per year in addition to variation in hours worked per week across workers.

The elasticities resulting from this exercise are in general lower but on the whole similar to than those in our baseline specification, with a Marshallian elasticity of 0.13 and a Hicksian elasticity of 0.42. The value of θ is larger than in our main results (at 2.3), and much less precisely estimated. The value of ϕ is essentially unchanged.

Table 22: MRS estimates using alternative hours definition

θ	2.29
	[−1.66,6.25]
ϕ	0.78***
	[0.53,1.04]
Wage elasticities at median	
Marshallian	0.13
	[−0.13,0.38]
Hicksian	0.42
	[−0.28,1.13]

N=50,895. *p<0.10, ** p<0.05, *** p<0.01

Online Appendix C: Probit results

Here we present results for the selection probit we run prior to estimating our MRS equation. Husband's earnings are strongly negatively correlated with participation.

Table 23: Selection Probit Results

Log earnings of husband	-0.164***	(0.007)
Husband employed	-1.929***	(0.064)
No. of Elderly HH members	0.023	(0.026)
Log family size	-0.110***	(0.022)
Wife: White	-0.015	(0.014)
Age	-0.056	(0.042)
Age^2	0.001	(0.001)
Age^3 /1000	0.003	(0.018)
Age^4 /10000	-0.003*	(0.001)
Has kids	-0.034	(0.018)
No. of kids aged 0-2	-0.515***	(0.014)
No. of kids aged 3-15	-0.167***	(0.008)
No. of kids aged 16-17	0.071***	(0.017)
North East	-0.004	(0.015)
Mid-West	0.119***	(0.014)
South	0.035**	(0.013)

N= 78,674. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Standard errors in parentheses. Additional controls for season and year dummies and cohort-education interactions.

Online Appendix D: Bootstrap procedure

We bootstrap standard errors and confidence intervals for both our MRS and Euler equations.

The two step Heckman-selection procedure for estimating the MRS coefficients can be bootstrapped in the standard way. Bootstrapping results for our Euler equation requires a slightly more complicated procedure however. This is because we aggregate our data into cohort groups and then implement an

IV procedure. Taking Z_t as a vector of exogenous variables, and X_t and Y_t as endogenous variables (with Y_t as our dependent variable) we can reformulate our approach as estimating the equations

$$\begin{aligned} X_t &= \Pi Z_t + v_t \\ Y_t &= X_t \beta + u_t \end{aligned}$$

where v_t is a vector of errors in our first stage. These can be thought of as economic shocks which may have a complicated structure. For instance they may be correlated across time for a given cohort, or may have an aggregate component which is correlated across cohorts for a given time period. Errors may also be correlated across the equations for different exogenous variables Z_t. We will wish to preserve these correlations when we implement our bootstrap procedure. In order to do this, we attempt to construct the variance-covariance matrix of the residuals v. Rather than filling in all possible cross-correlations in this matrix, we calculate the following moments for each cohort c, and equation i

$$\begin{aligned} &var(v^{i,c}) \\ &cov(v_t^{i,c}, v_{t-1}^{i,c}) \\ &cov(v_t^{i,c}, v_t^{j,c}) \\ &cov(v_t^{i,c}, v_t^{i,k}) \end{aligned}$$

Setting all other correlations to zero. Thus we impose for instance that there is zero correlation between $v_t^{i,c}$ and $v_{t-1}^{i,k}$. Unfortunately, there is no guarantee that this matrix will be postive definite. In our procedure we therefore apply weights to the non-zero elements of our 'off-diagonal' matrices - which give the covariances across different cohorts for the same equation - and to our 1st autocovariances for residuals for the same cohort and same equation. The weights we apply to these are the maximum that ensure the resulting matrix is positive defintite: in our case they are both set at 0.23.

Once we have this matrix we can Cholesky decompose it to obtain a vector of orthogonalised residuals

$$\Omega = vv' = \epsilon CC'\epsilon'$$

We then draw from the orthogonalised residuals, premultiply them by C and then add them to ΠZ_t to reconstruct the endogenous variables (including Y). We then reestimate our second stage equation to obtain a new set of estimates for β.

The values of Z_t in our case will depend on the results we obtain from our MRS equation, so in each iteration of our bootstrap we resample with replacement from from our disaggregated data, re-run the MRS equation, reaggregate to obtain the cohort averages which make up Z_t and then make a draw from our residuals.

Online Appendix E: Results for CES and additive separability

In this Appendix we discuss results for alternative specifications of our utility function. In particular we consider results from a standard CES utility function (where we impose that θ=ϕ), and one where we impose additive separability between consumption and leisure (i.e $\gamma = 0$).

Table 24 presents parameter estimates when we impose the restrictions implied by CES utility. Under this functional form for utility, we get a slightly larger value of ϕ and a much lower value of θ than we obtain from our preference specification (at 0.83 compared to 0.76 and 1.75 that we obtain for ϕ and θ respectively in Table 4). We also obtain a slighlty larger value of γ however (at 3.03 compared to 2.07 for our less restrictive utility function).

Table 24: Parameter values

	CES
ϕ	0.83 [0.66,0.97]
θ	0.83 [0.66,0.97]
γ	3.03 [0.64,4.27]

Taken together, the CES parameter estimates imply that utility is less concave in leisure, and hence that labour supply elasticities are greater. We show the elasticities implied by these estimates in Table 25. While Marshallian hours elasticities for the CES specification are only greater at the upper end of the distribution, the estimated Hicksian and Frisch hour elasticities are roughly 50% larger. The CES estimates also imply a more substantial degree of non-separability between consumption and leisure. The Frisch elasticity of consumption with respect to predictable wage increases has a median of around 0.4 compared to 0.05 from our main estimates. This reflects both a greater senstivity of the marginal utility of consumption to changes in leisure and the fact that leisure responses to given wage changes will in general be greater under these preferences. Finally we note that, the interest rate Frisch elasticity is much lower at the median than in our baseline specification, and, as a result of the imposed equality between θ and ϕ, identical to the interest rate Frisch for leisure.

Table 25 also shows Frisch elasticities for our preference specification in the case where we impose additive separability for preferences over consumption and leisure (that is we impose that $\gamma = 0$). This necessarily sets the Frisch consumption responses to wage changes to zero. It turns out that Frisch hours and leisure elasticities are very similar to those estimated when we allow for non-separability in our main specification. This reflects the fact that when, as we find, the parameters θ and ϕ are small and α large, then the numerator and denominator in formula for Frisch elasticities given in equation (25) will both be dominated by the term M_t. Consequently, the impact of small changes in γ will be limited.

When additive separability is imposed, the wage and interest rate Frisch elasticities are identical - a direct result of setting $u_{cl} = 0$ in expressions (26) and (27). The estimated Frisch elasticity of consumption with respect to the interest rate (now simply given by $-1/\phi$) also falls relative to our baseline results, from -1.19 to -1.31.

Table 25: Elasticities at Percentiles of Marshallian distribution: CES

	$\gamma = 0$		CES			
	Wage	Interest rate	Wage			Interest rate
	Frisch	Frisch	Marshallian	Hicksian	Frisch	Frisch
Hours worked						
10th	0.84	0.84	-0.24	0.48	1.08	0.83
	[0.22,3.14]	[0.22,3.14]	[-0.30,-0.11]	[0.41,0.60]	[0.97,1.45]	[0.62,1.31]
25th	0.83	0.83	-0.04	0.60	1.16	0.85
	[0.22,3.27]	[0.22,3.27]	[-0.13,0.12]	[0.52,0.76]	[1.06,1.54]	[0.64,1.35]
50th	0.90	0.90	0.21	0.77	1.33	0.93
	[0.24,3.59]	[0.24,3.59]	[0.10,0.42]	[0.66,0.99]	[1.24,1.75]	[0.71,1.48]
75th	1.04	1.04	0.54	1.04	1.66	1.13
	[0.28,4.31]	[0.28,4.31]	[0.39,0.82]	[0.89,1.32]	[1.54,2.19]	[0.86,1.80]
90th	1.98	1.98	1.11	1.62	2.71	1.89
	[0.57,6.96]	[0.57,6.96]	[0.88,1.55]	[1.39,2.07]	[2.49,3.65]	[1.47,3.03]
Leisure						
25th	-0.57	-0.57	-0.29	-0.57	-0.88	-0.59
	[-2.16,-0.16]	[-2.16,-0.16]	[-0.45,-0.21]	[-0.73,-0.49]	[-1.15,-0.81]	[-0.93,-0.45]
50th	-0.57	-0.57	-0.13	-0.48	-0.83	-0.58
	[-2.16,-0.16]	[-2.16,-0.16]	[-0.26,-0.06]	[-0.61,-0.41]	[-1.08,-0.76]	-0.91,-0.44]
75th	-0.57	-0.57	0.03	-0.40	-0.78	-0.57
	[-2.16,-0.16]	[-2.16,-0.16]	[-0.08,0.09]	[-0.50,-0.35]	[-1.04,-0.71]	[-0.90,-0.43]
Consumption						
25th	0.00	-1.31	0.91	0.63	0.32	-0.59
	[$-$,$-$]	[-1.81,-1.05]	[0.82,1.08]	[0.54,0.79]	[0.12,0.53]	[-0.93,-0.45]
50th	0.00	-1.31	1.07	0.72	0.37	-0.58
	[$-$,$-$]	[-1.81,-1.05]	[0.97,1.27]	[0.62,0.92]	[0.15,0.62]	-0.91,-0.44]
75th	0.00	-1.31	1.23	0.80	0.42	-0.57
	[$-$,$-$]	[-1.81,-1.05]	[1.12,1.44]	[0.68,1.02]	[0.17,0.70]	[-0.90,-0.43]

Elasticities are calculated as averages within 5 percent bands of the 10th, 25th, 50th and 75th and 90th percentiles of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

Online Appendix F: Returns to experience

In order to assess the ability of the model with returns to experience to account for female labour supply behaviour we provide several statistics beyond the targets of the calibration. First, analogously to Figure 2, Figure 4 shows life cycle profiles in the simulations and in the data. Second, Table 26 reports some additional statistics.

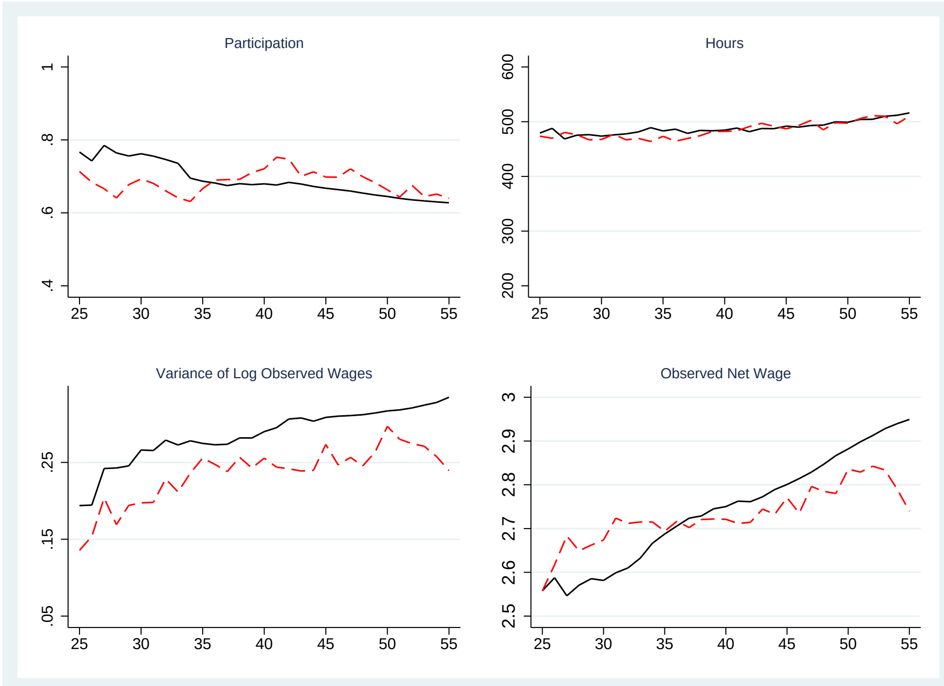
There are some differences between the model with returns to experience and the baseline we considered above. First, the median duration of spells out of the labour force is much longer: those who do exit, exit for long periods or do not return. This can be seen in the declining participation profiles at ages beyond 35. These patterns are not observed either in the data or in the baseline model. Second, very few women change their participation decisions. For example, the fraction of women who worked in all previous periods at the age of 52 is 57%, which compares to 40% in the economy without returns to experience. Third, the childcare cost that is needed here to keep women out of the labour market during childbearing is substantially higher because of the incentive to accumulate labour market experience. In particular the monetary fixed childcare cost is up to 76% of median earnings of a women aged 25 to 55. Finally, the size of the negative effect of husband earnings on the participation decision is slightly lower than in the baseline.

Table 26: Returns to Experience: Statistics on Hetereogeneity

	Data	Model
Participation Rate Mothers with Children Aged 3-17	0.682	0.672
Participation Rate Childless Women	0.730	0.724
Average Hours Worked 10th Percentile	260	277
Average Hours Worked 25th Percentile	455	400
Average Hours Worked 50th Percentile	520	518
Average Hours Worked 75th Percentile	520	595
Average Hours Worked 90th Percentile	624	648
Median Duration of Spells (years)		7
Wage 10th Percentile	8.17	10.60
Wage 50th Percentile	15.09	15.58
Wage 90th Percentile	29.45	31.71

Women without dependent children are women who have never had children and those whose children are over 17.

Figure 4: Life Cycle Profiles: Baseline Model (solid black line) versus Data (dashed red line)



Online Appendix G: Solution Method

Households have a finite horizon and so the model is solved numerically by backward recursion from the terminal period. At each age we solve the value function and optimal policy rule, given the current state variables and the solution to the value function in the next period. This approach is standard. The complication in our model arises from the combination of a discrete choice (to participate or not) and a continuous choice (over saving). This combination means that the value function will not necessarily be concave. We briefly describe in this appendix how we deal with this potential non-concavity.

In addition to age, there are four state variables in this problem: the asset stock, the permanent component of earnings of the husband, $v^m_{h,t}$, the permanent component of wife's wage, $v^f_{h,t}$,and the experience level of the wife. We discretise both earning and wage variables and the experience level, leaving the asset stock as the only continuous state variable. Since both permanent components of earnings are non-stationary, we are able to approximate this by a stationary, discrete process only because of the finite horizon of the process. We select the nodes to match the paths of the mean shock and the unconditional variance over the life-cycle. In particular, the unconditional variance of the permanent component must increase linearly with age, with the slope given by the conditional variance of the permanent shock. Our estimates of the wage variance are for annual shocks, but the model period is one quarter. We reconcile this difference by imposing that each quarter an individual receives a productivity shock with probability 0.25, and this implies that productivity shocks occur on average once a year.

Value functions are increasing in assets A_t but they are not necessarily concave, even if we condition on labour market status in t. The non-concavity arises because of changes in labor market status in future periods: the slope of the value function is given by the marginal utility of consumption, but this is not monotonic in the asset stock because consumption can decline as assets increase and expected labour market status in future periods changes. By contrast, in Danforth (1979) employment is an absorbing state and so the conditional value function will be concave. Under certainty, the number of kinks in the conditional value function is given by the number of periods of life remaining. If there is enough uncertainty, then changes in work status in the future will be smoothed out leaving the expected value function concave: whether or not an individual will work in $t+1$ at a given A_t depends on the realisation of shocks in $t+1$. Using uncertainty to avoid non-concavities is analogous to the use of lotteries elsewhere in the literature.

The choice of participation status in t is determined by the maximum of the conditional value functions in t. In our solution, we impose and check restrictions on this participation choice. In particular, we use the restriction that the participation decision switches only once as assets increase, conditional on permanent earnings and experience. When this restriction holds, it allows us to interpolate behaviour across the asset grid without losing our ability to determine participation status. We therefore define a reservation asset stock to separate the value function and the choice of consumption made when participating from the value function and choice of consumption made when not participating. There are some regions of the state space where individuals are numerically indifferent between working and not-working. Since we solve the model by value function iteration, it does not matter which conditional value function we use in these regions.

In solving the maximisation problem at a given point in the state space, we use a simple golden search method. Note that in addition to the optimal total expenditure, the optimal amount of leisure is computed in each period by solving the MRS condition. We solve the model and do the calibration assuming this process is appropriate and assuming there is a unique reservation asset stock for each point in the state space, and th en check ex-post.

There are no non-concavities due to borrowing constraints in our model because the only borrowing constraint is generated by the no-bankruptcy condition which is in effect enforced by having infinite marginal utility of consumption at zero consumption.