Aggregating Elasticities: Intensive and Extensive Margins of Women's Labour Supply

Online Appendix

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

[*]UCL, IFS and NBER
[†]IFS and UCL
[‡]University of Cambridge and IFS
[§]Universidad de Cantabria

Online Appendix A (to section 2): Derivation of Elasticities

Marshallian and Hicksian Elasticities

$$y_t = \left(A_{h,t} + y^m_{h,t} - F\left(a_{h,t}\right) P_{h,t}\right) - \frac{A_{h,t+1}}{1+r_{t+1}} \tag{21}$$

$$c_t + w_t l_t = y_t + w_t L \tag{22}$$

$$w_{h,t} = \frac{u_{l_{h,t}}}{u_{c_{h,t}}} = \alpha_{h,t} \frac{l_{h,t}^{-\theta}}{c_{h,t}^{-\phi}} \tag{23}$$

Taking the derivative of the budget constraint and the MRS equation and stacking them gives a matrix equation:

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

This can be inverted to give the Marshallian elasticities in the main text (equation 14):

$$\varepsilon_c^M = \frac{\partial \ln c}{\partial \ln w} = \frac{\theta w\left(L-l\right) + wl}{\theta c + \phi wl} \tag{24}$$

$$\varepsilon_l^M = \frac{\partial \ln l}{\partial \ln w} = \frac{\phi w\left(L-l\right) - c}{\theta c + \phi wl} \tag{25}$$

$$\varepsilon_h^M = \frac{\partial \ln h}{\partial \ln w} = -\left(\frac{\phi w\left(L-l\right) - c}{\theta c + \phi wl}\right) \frac{l}{L-l} \tag{26}$$

To calculate the Hicksian elasticities, we first calculate the income elasticities by differentiating the MRS equation and the budget constraint with respect to income:

$$\varepsilon_c^y = \frac{\partial \ln c}{\partial \ln y} = \frac{\theta y}{\theta c + \phi wl} \tag{27}$$

$$\varepsilon_l^y = \frac{\partial \ln l}{\partial \ln y} = \frac{\phi y}{\theta c + \phi wl} \tag{28}$$

The income elasticity and the marshallian elasticity are then used to calculate the Hicksian elasticity using the Slutsky equation:

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

Frisch Elasticities

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:

$$u_c\left(c,l\right) = DM^{-\gamma}c^{-\phi} \tag{29}$$

$$u_l\left(c,l\right) = D\alpha M^{-\gamma}l^{-\theta} \tag{30}$$

$$u_{cl}\left(c,l\right) = (-\gamma)DM^{-\gamma-1}\alpha c^{-\phi}l^{-\theta} \tag{31}$$

$$u_{ll}\left(c,l\right) = (-\gamma)\frac{u_l\left(c,l\right)}{\alpha M}l^{-\theta} - u_l\left(c,l\right)\theta l^{-1} \tag{32}$$

$$u_{cc}\left(c,l\right) = (-\gamma)\frac{u_c\left(c,l\right)}{M}c^{-\phi} - u_c\left(c,l\right)\phi c^{-1} \tag{33}$$

Finally, note that:

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

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:

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

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

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

$$\varepsilon_c^F = \frac{w}{c}\frac{\partial c}{\partial w} = -\frac{u_c u_{cl}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{c} \tag{35}$$

$$\varepsilon_l^F = \frac{w}{l}\frac{\partial l}{\partial w} = \frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{l} \tag{36}$$

$$\varepsilon_h^F = \frac{w}{L-l}\frac{\partial (L-l)}{\partial l}\frac{\partial l}{\partial w} = -\frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2}\frac{w}{L-l} = -\varepsilon_l\frac{l}{L-l} \tag{37}$$

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

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

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

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

$$\varepsilon_c^{F_R} = \frac{(1+R_{t+1})}{c} \frac{\partial c}{\partial(1+R_{t+1})} = \frac{u_c u_{ll} - u_l u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_c u_{ll} - u_l u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)} \tag{38}$$

$$\varepsilon_l^{F_R} = \frac{(1+R_{t+1})}{l} \frac{\partial l}{\partial(1+R_{t+1})} = \frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_l u_{cc} - u_c u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)} \tag{39}$$

$$\varepsilon_h^{F_R} = \frac{(1+R_{t+1})}{L-l} \frac{\partial(L-l)}{\partial l} \frac{\partial l}{\partial(1+R_{t+1})} = -\frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{L-l} = -\varepsilon_l^{F_R} \frac{l}{L-l} \tag{40}$$

Online Appendix B (to section 3): Estimation Strategy and Solution Method

Fuller's estimator

When estimating the parameters that determine the MRS or those that enter the Euler equation, we use first order conditions to derive restrictions on the data to identify structural parameters. 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

$$E\left[h(X;\theta)\mathcal{Z}\right] = 0 \tag{41}$$

where $h(\cdot)$ is a function of data X and parameters, θ, and is linear in the vector of parameters. The vector $\mathcal{Z}$ contains 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 (41).

In equation (41), one needs to normalise one of the parameters to 1. In the context of the MRS equation (17), for example, we 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. This estimator is a modified version of LIML with an adjustment that is designed to ensure that it has finite moments. Roughly speaking, it can be thought of as a compromise between LIML and 2SLS (being closer to LIML when the sample size is large relative to the number of instruments). 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 section 3, 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).[1] 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.

[1] 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.

Euler Equation Estimation with Repeated Cross-Sections

We estimate the intertemporal parameters using the Euler equation. We need a long time series because, 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), Attanasio (1999), or Attanasio and Weber (2010). We also need to assume that the lagged variables used as instruments are uncorrelated with the innovations to the taste shifters $\Delta\zeta_{h,t+1}$. This is trivially true if taste shifters are constant over time or if they are random walks. We maintain one of these two assumptions, a hypothesis that we can in part test by checking over-identifying restrictions.

We estimate equation (20) using the Consumer Expenditure Survey (CEX). Although the CEX covers many years, each household is only observed for a few quarters and so we use a synthetic cohort approach (see Browning et al. (1985)): we aggregate equation (20) over groups with constant membership and follow the average behaviour of the variables of interest (or their non-linear transformation) for such groups. A time series of quarterly 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 married couples in 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.[2]

Having identified groups, we aggregate equation (20) to be estimated across group g households. For this approach to work, however, it is necessary that the equation to be estimated is linear in parameters, which would be the case if $M_{h,t}$ were observable. 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 $M_{g,t}$. On the other hand, the parameters that determine $M_{h,t}$ can be consistently estimated using the MRS conditions as discussed in section 3.1.[3] These estimates can be used to construct consistent estimates of $M_{h,t}$, which can be aggregated across households to give $M_{g,t}$.

We can obtain consistent estimates of the grouped variables from the time series of cross sections,

[2]If credit constraints are binding, the Euler equation will not be holding as an equality. The youngest 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 may be difficult to capture with the taste shifters that we have been considering.

[3]$M_{h,t}$ includes $\chi_{h,t}$ which is unobserved. However, since it is the residual from the MRS equation, it can be included in the calculation of $\alpha_{h,t}$ that is needed to calculate $M_{h,t}$.

giving the group average log-linear Euler equation:

$$\widetilde{\eta}_{g,t+1} = \overline{\kappa} + \ln\beta + \ln(1+r_{t+1}) - \phi\Delta\overline{\ln c_{g,t+1}} + -\gamma\Delta\ln(\overline{\widehat{M}_{g,t+1}}) + \varphi\Delta\overline{P_{g,t+1}} + \pi\Delta\overline{z_{g,t+1}} \tag{42}$$

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: (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 (42) are observable. We can therefore use this equation to estimate the parameters of interest. However, the instruments need to be uncorrelated with $\widetilde{\eta}_{g,t+1}$.[4] 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 have effects that correlate across different 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 matrix of $\widetilde{\eta}_{g,t+1}$ from 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 follow a different and, as far as we know, novel approach, based on bootstrapping our sample, with a structure consistent with the basic assumptions of our model. We describe the bootstrapping procedure in detail in the next subsection.

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

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

[4]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, because of 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).

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 positive 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 definite: 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.

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. An alternative would be to follow the method in Iskhakov et al. (2017).

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 labour 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 then 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.

Finally, we include here the value functions of the household problem. In each period, if the woman chooses to participate, the value function is given by

$$V_{h,t}^{1}\left(A_{h,t},v_{h,t}\right)= \max_{c_{h,t},l_{h,t}}\left\{u\left(c_{h,t},l_{h,t},P_{h,t}=1\right)+\ \beta E_t\left[\max\left\{\begin{array}{c} V_{h,t+1}^{0}\left(A_{h,t+1},v_{h,t+1}\right) \\ V_{h,t+1}^{1}\left(A_{h,t+1},v_{h,t+1}\right)\end{array}\right\}\right]\right\} \tag{43}$$

Note that the state variable $v_{h,t}$ is a vector containing the woman and the man's productivity type. If she chooses not to participate, the value function is given by,

$$V_{h,t}^{0}\left(A_{h,t},v_{h,t}\right)= \max_{c_{h,t}}\left\{u\left(c_{h,t},P_{h,t}=0\right)+\ \beta E_t\left[\max\left\{\begin{array}{c} V_{h,t+1}^{0}\left(A_{h,t+1},v_{h,t+1}\right) \\ V_{h,t+1}^{1}\left(A_{h,t+1},v_{h,t+1}\right)\end{array}\right\}\right]\right\} \tag{44}$$

The decision of whether or not to participate in period t is determined by comparing $V_{h,t}^{0}\left(A_{h,t},v_{h,t}\right)$ and $V_{h,t}^{1}\left(A_{h,t},v_{h,t}\right)$. The participation choice, the hours choice and the consumption choice in t determines the endogenous state variable (assets) at the start of the next period.

Online Appendix C (to section 4): Data Sources and Descriptive Statistics

As discussed in the paper, most of the data are from the CEX. One important exception are the data on the real interest rate. We define this variable as the 3 month T-Bill rate (on a quarterly basis) minus the rate of growth in the CPI. The source for the T-Bill rate is from the St Louis Fed (https://fred.stlouisfed.org/series/TB3MS).

In for Table 12 presents descriptive statistics at the individual level using data from three particular years (1980, 1995 and 2012). Married women have seen large changes in their wages, hours and patterns of employment over our sample period. Employment rates increased from 60% in 1980 to 69.8% in 1995 before falling back to 61.9% in 2012.

Table 12: Descriptive statistics for married women, 1980, 1995 and 2012

		1980	1995	2012
Demographics	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
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	14.62
	Some college	16.62	16.41	17.28
	Degree or higher	19.30	22.26	23.20
% Employed	All workers	60.0	69.8	61.9
	% Workers part-time	28.4	23.7	20.6
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.

Table 12 also shows wage levels over the three years. Average real wages increased over this period, though with marked differences across different education groups. 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.

Online Appendix D (to section 5)

Alternative methods of estimating the MRS

In this appendix we discuss results from alternative MRS specifications. For comparison with later results, we present a more complete set of parameter estimates from our baseline MRS specification in Table 14. First, 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 13: 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.

Estimation method and normalisation

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

$$\ln w_{h,t} = \psi_0 + \psi z_{h,t} - \theta \ln l_{h,t} + \phi \ln c_{h,t} + \upsilon_{h,t} \tag{45}$$

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 15 shows results from estimating ϕ and θ using GMM under the three different possible normalisations.

Table 14: 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)	[-0.002,0.001]
Age^3/1000	-0.01	(0.01)	[-0.03,0.01]
Age^4/10000	0.002**	0.0007	[0.0002,0.003]
North East	0.01	(0.03)	[-0.02,0.08]
Mid West	-0.05**	(0.01)	[-0.07,-0.02]
South	-0.11***	(0.02)	[-0.18,-0.09]
White	-0.04	(0.03)	[-0.09,0.04]
No. elderly HH members	0.02	(0.02)	[-0.02,0.05]
ln($famsize$)	-0.32***	(0.037)	[-0.38,-0.23]
Has kids	0.07***	(0.021)	[0.04, 0.12]
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)	[-1.19,18.52]
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.

We include results form our baseline specification in the first column. The implied parameter estimates and elasticities vary a great deal across these different approaches. When wages are selected as the left-hand side variable, elasticites are relatively large. 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 (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

Table 15: MRS Estimates using GMM

	Fuller	GMM		
Dependent variable:	Wages	Wages	Leisure	Consumption
Parameters				
θ	1.75** [0.34,5.12]	0.46* [−0.04,0.61]	13.8 [−120.13,186.11]	0.13 [−0.54,0.58]
ϕ	0.76*** [0.55,0.95]	0.61*** [0.48,0.66]	0.17 [−3.44,2.78]	1.38*** [1.24,1.74]
Wage elasticities at median				
Marshallian	0.18 [0.05,0.38]	0.55 [0.50,1.16]	0.09 [0.00,0.12]	−0.17 [−0.41,−0.08]
Hicksian	0.54 [0.27,1.29]	1.19 [1.10,2.25]	0.11 [−0.01,0.14]	0.77 [0.59,1.10]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 3. 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.

bias properties in such circumstances (Staiger and Stock, 1997; Nelson and Startz, 1990). Using the notation from Davidson and MacKinnon (2004), for the case where

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

where Z is a matrix of exogenous variables, Y a matrix of endogenous variables, and $W = [Z, W_1]$ (with W_1 being a matrix of instruments). Matrices X and W are $n \times k$ and $n \times l$ respectively (with $l \geq k$). In general, so-called k-class estimators such as OLS, 2SLS, and LIML can be written in the form

$$\hat{\beta}^{\text{LIML}} = (X'(I - kM_W)X)^{-1}X'(I - kM_W)y \tag{46}$$

where $M_W = I - W(W'W)W'$. In the case of OLS $k = 0$, and in the case of 2SLS $k = 1$. In the case of LIML we use

$$k = k_{\text{LIML}} = \frac{(y - Y\beta_2)'M_Z(y - Y\beta 2)}{(y - Y\beta_2)'M_W(y - Y\beta 2)} \tag{47}$$

While LIML is often found to have better bias properties than 2SLS, it has long been recognised that conventional normalisations of LIML do not have finite moments (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 in equation (46) with

$$k_{\text{Fuller}} = k_{\text{LIML}} - \frac{\lambda}{(n-k)} \tag{48}$$

where λ here is a parameter chosen by the researcher, to obtain a value for $\hat{\beta}^{\text{Fuller}}$. We choose a value of one for this as suggested by Davidson and MacKinnon (2004) as it yields estimates that are approximately unbiased. The resulting estimator is guaranteed to have bounded moments in finite samples Fuller (1977). Since the adjustment to LIML is smaller when $(n-k)$ is large, the Fuller estimator will be closer to LIML when sample sizes are large relative to the number of instruments. In our case, the Fuller estimator can be thought of as a compromise between 2SLS and LIML, as it adjusts the value of k we use downwards slightly towards one.

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 16 shows. Both the elasticity and parameter estimates obtained using alternative normalisations of the Fuller estimator are very similar to our baseline results.

Table 16: MRS Estimates with Different Dependent Variables

	Dependent variable		
	Wages	Leisure	Consumption
Parameters			
θ	1.75**	1.84*	1.75*
	[0.34,5.12]	[−0.43,5.38]	[−0.00,4.60]
ϕ	0.76***	0.76***	0.77***
	[0.55,0.95]	[0.53,0.95]	[0.58,0.95]
Wage elasticities at median			
Marshallian	0.18	0.17	0.18
	[0.05,0.38]	[0.06,0.37]	[0.07,0.42]
Hicksian	0.54	0.53	0.54
	[0.24,1.07]	[0.23,0.95]	[0.27,1.29]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 3. 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 17 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, in both cases using a *full* set of cohort-education-year interactions as used in Blundell et al. (1998). This approach 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 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 of 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 when using the alternative instrument set.

Table 17: MRS Estimates using Alternative Instruments

	Fuller	GMM
Parameters		
θ	1.93 [−9.09,11.58]	0.08 [−0.20,0.19]
ϕ	0.76*** [0.42,1.03]	0.52*** [0.41,0.52]
Wage elasticities at median		
Marshallian	0.17 [−0.84,1.13]	1.08 [1.04,2.29]
Hicksian	0.51 [−1.91,2.65]	1.97 [1.82,3.85]

N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. We use a full set of cohort-education-year dummies as instruments, following Blundell, Duncan and Meghir (1998). Controls as in Table 3. 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 18 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. Excluding these observations, 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 18: MRS Estimates using alternative samples/hours measures

	Exc. 40 hours (1)	Exc. <20 hours (2)	Born 1925-1965 (3)	Ann. hours (4)
Parameters				
θ	1.52 [−3.16,5.69]	2.81 [−2.48,9.81]	2.08*** [0.68,4.66]	2.30* [−1.30,7.00]
ϕ	0.42* [−0.05,0.92]	0.76*** [0.33,1.03]	0.56*** [0.42,0.82]	0.78*** [0.53,1.01]
Wage elasticities at median				
Marshallian	0.45 [−0.45,1.41]	0.13 [0.02,0.30]	0.27 [0.09,0.62]	0.13 [−0.09,0.41]
Hicksian	0.72 [−1.29,2.57]	0.39 [−0.14,1.43]	0.53 [0.27,1.09]	0.42 [−0.25,1.24]
N	26,060	47,743	39,057	50,895

*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 Column (4) of Table 18 we consider how elasticity estimates are affected when we use an alternative measure of hours of leisure. The measure we use here is

$$\text{leisure} = \frac{5200 - \text{hours per week} \times \text{weeks worked per year}}{52} \tag{49}$$

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 2.30), and much less precisely estimated. The value of ϕ is essentially unchanged.

External Parameters for the Calibration

Table 19 reports the complete set of estimated and external parameters used in the calibration. The first panel reports the estimated parameters from Tables 3 and 4 above. The second panel reports parameters which come from external sources.

Table 19: 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)	$\sigma_{\xi_0^m}$	0.54
Length of Life (in years)	T	50
Length of Working Life (in years)	R	40

Online Appendix E (to section 6): 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 20 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 3). We also obtain a slightly larger value of γ however (at 3.04 compared to 2.07 for our less restrictive utility function).

Table 20: Parameter values

	CES
ϕ	0.83 [0.66,0.97]
θ	0.83 [0.66,0.97]
γ	3.04 [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 21. 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 sensitivity 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 at the median is much lower than in our baseline specification.

Table 21 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 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 formulae for Frisch elasticities given in equations (35) and (36) will be dominated by the term M_t. Consequently, the impact of small changes in γ will be limited.

When additive separability is imposed, the Frisch elasticity is identical - a direct result of setting $u_{cl} = 0$ in expressions (38) and (39). 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 a median value of -1.19 in our baseline results to -1.31.

Table 21: Elasticities at Percentiles of Marshallian distribution: CES

	$\gamma = 0$		CES			
	Wage	Interest rate	Wage			Interest rate
	Frisch	Frisch	Marshallian	Hicksian	Frisch	Frisch
	Hours worked		*Hours worked*			
10th	0.84 [0.23,3.17]	0.84 [0.23,3.17]	−0.24 [−0.30,−0.11]	0.48 [0.41,0.60]	1.08 [0.97,1.47]	0.83 [0.63,1.32]
25th	0.83 [0.23,3.15]	0.83 [0.23,3.15]	−0.04 [−0.13,0.12]	0.60 [0.51,0.76]	1.16 [1.06,1.56]	0.85 [0.65,1.35]
50th	0.90 [0.25,3.40]	0.90 [0.25,3.40]	0.21 [0.10,0.42]	0.77 [0.67,0.98]	1.33 [1.23,1.75]	0.93 [0.71,1.47]
75th	1.04 [0.29,3.93]	1.04 [0.29,3.93]	0.54 [0.39,0.82]	1.04 [0.89,1.32]	1.66 [1.53,2.17]	1.13 [0.86,1.78]
90th	1.98 [0.55,7.50]	1.98 [0.55,7.50]	1.11 [0.88,1.55]	1.62 [1.39,2.06]	2.71 [2.51,3.57]	1.89 [1.44,2.99]
	Consumption		*Consumption*			
25th	0.00 [−,−]	−1.31 [−1.81,−1.05]	0.91 [0.82,1.08]	0.63 [0.54,0.80]	0.32 [0.12,0.53]	−0.59 [−0.93,−0.45]
50th	0.00 [−,−]	−1.31 [−1.81,−1.05]	1.07 [0.97,1.27]	0.72 [0.62,0.91]	0.37 [0.15,0.62]	−0.58 −0.91,−0.44]
75th	0.00 [−,−]	−1.31 [−1.81,−1.05]	1.23 [1.12,1.44]	0.80 [0.68,1.01]	0.42 [0.17,0.69s]	−0.57 [−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

We recalibrate parameter values: 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.[5] As in the baseline, we identify these parameters by targeting participation rate of women, 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 (we take this figure from Attanasio et al. (2008)). We report the calibrated parameters in Table 22 and compare them to the baseline. In the context of returns to experience, where there is a strong incentive to work to reap future returns, a much larger childcare cost is required in order to reduce participation and match participation statistics.

Analogously to Figure 2, Figure 4 shows life-cycle profiles in the simulations and in the data; and Table 23 reports additional statistics on the distribution of hours and of wages. There are some differences between the model with returns to experience and the baseline. First, there is a decline in the 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 fix childcare cost is up to 76% of median earnings of a women aged 25 to 55.

Response to Temporary Wage Changes

In Table 24, 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 close to zero and, as a result, the aggregate elasticity is about half of the one in the baseline economy (reproduced in the final column). In the return to experience economy, there is a strong incentive to participate to obtain the return to experience. The larger childcare cost of participating that is estimated in this economy alongside the strong incentive to participate implies that changes in the current wage makes 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. Our results here are in line with Imai and Keane (2004) who argue that the response of labour supply to transitory changes in wages may be mitigated when there are returns to experience. Our results show

[5] 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 22: 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	0.78	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
Women's Human Capital Tech	ν	0.003	-
Discount Factor (annualized)	β	0.99	0.99
Depreciation rate	δ	0.017	-

Targets	Data	Ret to Exp	Baseline
Weekly hours worked	37.2	37.5	37.2
Participation Rate	0.684	0.690	0.679
Participation Rate of Mothers	0.538	0.544	0.546
Observed Wage Gender Gap	0.720	0.716	0.727
Observed Variance Wage Growth (Women)	0.004	0.005	0.004
Observed Initial Variance of Wages (Women)	0.14	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

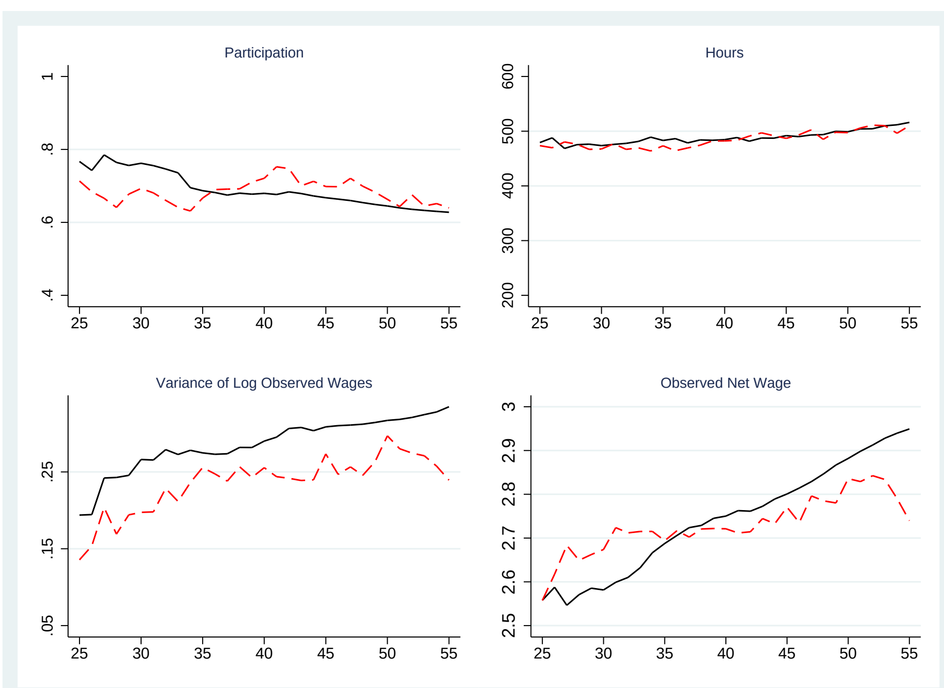
Statistics for women born in the 1950s and aged 25 to 55. Wage growth and depreciation rate are annual.

Table 23: Returns to Experience: Statistics on Hetereogeneity

	Data	Model
Participation Rate Mothers with Children Aged 3-17	0.682	0.672
Participation Rate Childless Women	0.755	0.724
Average Hours Worked 10th Percentile	20	21
Average Hours Worked 25th Percentile	35	31
Average Hours Worked 50th Percentile	40	40
Average Hours Worked 75th Percentile	40	46
Average Hours Worked 90th Percentile	48	50
Wage 10th Percentile	8.16	7.43
Wage 50th Percentile	15.05	15.58
Wage 90th Percentile	29.23	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)



that the response of the participation margin to a transitory anticipated change in the wage (for given preferences on the intensive margin) may be very different depending on the assumption that is made about the nature of the wage growth over the life-cycle (exogenous or endogenous). The extra wage that is provided by an anticipated increase in the wage in a particular period is a small fraction of the total return to participate in that period (in particular at early ages) and then it has a small impact on the participation decision.

Table 24: Returns to experience: Frisch Changes

	Extensive Response	Intensive Elasticity 25th	50th	75th	Agg Hours Elasticity	Baseline
25-29	0.02	0.65	0.81	1.15	0.91	1.85
30-34	0.04	0.63	0.79	1.17	0.91	1.48
35-39	0.03	0.63	0.78	1.17	0.90	1.45
40-44	0.03	0.61	0.79	1.19	0.89	1.35
45-49	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 percentage point change in participation in response to a 1% increase in the wage. The aggregate hours elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

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 core of the paper.

One possibility, of course, is that returns to tenure are important for some occupations and/or skill 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.[6]

Life-Cycle Responses to Changes in Wage Profiles

Finally, in Table 25 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

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

the extensive and the intensive margin is very similar in the economy with and without returns to experience.

Table 25: Labour supply changes, Marshallian

	Extensive Response	Intensive Elasticity 25th	50th	75th	Agg Hours Elasticity
Ret to experience	0.53	0.25	0.40	0.77	0.99
Baseline	0.51	0.28	0.42	0.67	0.91

The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate hours elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

Online Appendix G (to section 3): Selection Correction

In this Appendix we consider an extension of the full model that we calibrate in section 5.3. We allow for taste shocks χ and ζ in the utility function. We solve and simulate this economy and explore the ability of our empirical strategy in section 3 to recover the preference parameter values that are assumed in the simulations ($\phi = 0.76$ and $\theta = 1.75$). We discretise both χ and ζ.

In the first column of Table 26 we report the OLS estimates of the MRS equation using simulated data. We estimate $\phi = 0.52$ and $\theta = 1.97$. These are clearly biased with respect to the assumed parameter values. As discussed in section 3, there are two reasons for the bias. First, regressors are endogenous since consumption and leisure are correlated with the error term in the MRS equation, and, second, there is non-random self-selection of women into the labour market. In order to address the first issue we solve and simulate the economy for several cohorts of women that differ in the average wage they face. We use the variation in wages across cohorts to instrument consumption and leisure in the MRS equation and we provide the estimates in the second column of Table 26. Estimated parameter values are still biased with respect to the assumed parameter values. In order to address the consequences of non-random selection we estimate a probit of the participation decision in which we include as exclusion restrictions a cohort dummy and the log of husband earnings (the same as what we used in the data). We then include the selection correction terms as regressors of the MRS equation and report the results in the third column of the Table. In this case estimated parameter values are very close to the assumed parameter values.

Table 26: MRS estimates with simulated data

	OLS	IV	IV+Selection Correction
ϕ	0.52***	0.61***	0.76***
	(0.000130)	(0.00229)	(0.0666)
θ	1.97***	2.20***	1.71*
	(0.000561)	(0.00528)	(0.902)

Standard errors in parentheses.$*p < 0.10$, $**p < 0.05$,$***p < 0.01$

References

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., M. Costa, C. Meghir, and J. Shaw (2016): "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.

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.

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.

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.

Hahn, J. and J. Hausman (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.

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

Iskhakov, F., T. H. Jørgensen, J. Rust, and B. Schjerning (2017): "The Endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks," *Quantitative Economics*, 8, 317–365.

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

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.

Nelson, C. R. and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instumental Variables Estimator," *Econometrica*, 58, 967–976.

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.