

IZA DP No. 6098

**A Continuous Labour Supply Model in Microsimulation:  
A Life-Cycle Modelling Approach with  
Heterogeneity and Uncertainty Extension**

Jinjing Li  
Denisa Maria Sologon

October 2011

# **A Continuous Labour Supply Model in Microsimulation: A Life-Cycle Modelling Approach with Heterogeneity and Uncertainty Extension**

**Jinjing Li**

*Maastricht University / UNU-MERIT*

**Denisa Maria Sologon**

*CEPS/INSTEAD,*

*Maastricht University and IZA*

Discussion Paper No. 6098

October 2011

IZA

P.O. Box 7240

53072 Bonn

Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **A Continuous Labour Supply Model in Microsimulation: A Life-Cycle Modelling Approach with Heterogeneity and Uncertainty Extension\***

This paper advances a structural inter-temporal model of labour supply that is able to simulate the dynamics of labour supply in a continuous setting and to circumvent two main drawbacks of most of the existing models. The first limitation is the inability to incorporate individual heterogeneity as every agent is sharing the same parameters of the utility function. The second one is the strong assumption that individuals make decisions in a world of perfect certainty. Essentially, this paper offers an extension of marginal-utility-of-wealth-constant labour supply functions known as “Frisch functions” under certainty and uncertainty with homogenous and heterogeneous preferences. Two alternative models are proposed for capturing individual heterogeneity. First, a “fixed effect vector decomposition” model, which allows the individual specific effects to be correlated with the explanatory variables included in the labour supply model, and second, a mixed fixed and random coefficient model, which incorporates a higher degree of individual heterogeneity by specifying individual coefficients. Uncertainty is controlled for by introducing an expectation correction into the model. The validation of each simulation model is realized in comparison with the standard Heckman model. The lifetime models based on the fixed effect vector decomposition yield the most stable and unbiased simulation results, both under certainty and uncertainty. Due to its improved accuracy and stability, this lifetime labour supply model is particularly suitable for enhancing the performance of the pension models, thus providing a better reference for policymaking.

JEL Classification: C20, D90, J22

Keywords: lifetime labour supply, dynamic microsimulation

Corresponding author:

Denisa Maria Sologon  
CEPS/INSTEAD  
3, avenue de la Fonte  
4364 Esch-sur-Alzette  
Luxembourg  
E-mail: [Denisa.Sologon@ceps.lu](mailto:Denisa.Sologon@ceps.lu)

---

\* This research is part of the “Earnings Dynamics and Microsimulation” project supported by the Luxembourg “Fonds National de la Recherche” through an AFR grant (PDR no. 893613) under the Marie Curie Actions of the European Commission (FP7-COFUND). The authors would like to thank Cathal O’Donoghue for his helpful comments and advice.

## I. Introduction

The empirical literature on labour supply has gained an increasing interest over the past decades. The continuous approach in the tradition of Burtles and Hausman (1978) has been complemented by an approach which focuses on a discrete choice specification, mainly inspired by Van Soest (1995). Most studies focus on estimating and simulating labour supply in a static context, whereas only a few focus on the dynamics of labour supply (Blundell and MaCurdy, 1999).

Some of the existing models for estimating and/or simulating labour supply in a static or dynamic setting lack the robustness of economic theory. Other models are based on specifying utility functions used to derive consistent labour supply functions, which is fundamental for projecting the labour supply behaviour. The latter models, however, suffer from two major shortcomings. First, as utility is not observed, many assumptions are assumed in order to estimate the parameters of the utility function. This becomes less robust when more items/behaviours are introduced into the utility function. Secondly, this approach does not incorporate individual heterogeneity as every agent shares the same parameters.

Some models deal with the heterogeneity problem by capturing the individual specific effects using error components models or by estimating random-coefficients models. The individual specific effect is used in the context of the fixed or random effects estimations, where the main issue concerns the choice between fixed and random effects and whether the individual specific effects can be assumed to be independent of the explanatory variables included in the model. Assuming that the individual component is correlated with the explanatory variables triggers many problems when undertaking a simulation. The estimated coefficients cannot be used to generate a conditional prediction of individual earnings without specifying the joint process determining the individual specific effects and the explanatory variables (Pudney, 1992). The impracticality of this option, together with the fact that the fixed-effect specification cannot accommodate covariates that are constant over time, constrained most studies to maintain the assumption of a zero correlation between the individual specific effects and the other covariates, a rather strong and improbable assumption. The main drawback of the error components models is that they provide less stable simulations due to the stochastic components, which affects the reproducibility of the results.

One way to incorporate the heterogeneity effects is to use random coefficient models. Provided that heterogeneity is present in the empirical models of labour supply, the application of random coefficients models is necessary to avoid biased estimates. The main drawback of these models, however, is their high computational cost. Given this limitation, existing studies suggest that if heterogeneity is non-existent or the bias is insignificant, the standard fixed coefficient models represent the optimal choice (Haan, 2004). Several studies estimating discrete choice labour supply models compare the fixed and random specifications and find no significant differences between the results (van Soest, 1995; Duncan and MacCrae, 1999; Haan, 2004). The estimation of continuous labour supply models using the random coefficient specification is limited.

This chapter proposes a structural inter-temporal model of labour supply, which estimates and predicts the dynamics of labour supply in a continuous setting. It aims to capture the individual heterogeneity to a larger extent than the existing labour supply models while maintaining the consistency with the lifetime economic theory. The model is estimated using both a transformed fixed effect specification that circumvents the

standard problems mentioned above, and a random coefficient specification. Additionally, the model incorporates uncertainties regarding future wages to further explore the heterogeneities.

## II. Theoretical Background

Most of the empirical work on labour supply is based on a static, within-period framework, where the labour supply decision rule refers to one period. Typically, the annual hours of work are regressed on the current hourly wage rate and some measure of property income. Labour supply responses are estimated using the standard labour supply models, which ignore the importance of future wages on current hours supplied. This yields a wage coefficient that confuses the labour supply response to three types of wage changes: “evolutionary wage changes” arising from movements along the lifetime wage profile, “parametric wage changes” arising from shifts in the wage profile, and those arising from changes in the profile slopes. As a result, the wage coefficients reported in many empirical studies have no economic interpretation and are not useful for policy evaluations (MaCurdy, 1980).

The elasticity derived from the static specification can be placed in an inter-temporal setting, but is economically meaningful only under the strong assumption of either complete myopia or perfectly constrained capital markets. These assumptions imply that it is impossible to transfer capital across periods. In this situation, the coefficient on wage represents the uncompensated substitution elasticity given income, equivalent to the Marshallian wage elasticity in the static model (Blundell and MaCurdy, 1998).

There are three types of substitution elasticities relevant for predicting the response of hours of work to changes in the wage rate: the inter-temporal elasticity, which determines the labour supply wage changes resulting from life-cycle wage growth and movements over a perfectly foreseen business cycle, the uncompensated elasticity, which determines the labour supply response to shifts in wage profiles whilst holding the marginal utility of wealth constant, and the compensated elasticity, which can be used to predict the differences in hours of work across consumers with different wage profiles but an identical lifetime utility constant. In order to estimate meaningful behavioural parameters for the inter-temporal and the uncompensated substitution effects, it is crucial to recognize that individuals make their labour supply decisions within a life-cycle framework. Moreover, formulating a model that relies on the economics of lifetime behaviour leads to a better understanding of consumer behaviour (Blundell and MaCurdy, 1998; MaCurdy, 1980).

The acknowledgement that labour supply is part of a lifetime decision process is realized by the multi-period models of labour supply. In the context of this chapter, special attention is devoted to the Frisch labour supply functions, which estimate the effect of labour supply whilst holding the marginal utility of wealth constant. These functions are useful when analysing life-cycle maximization problems. The Frisch labour supply functions represent a third type of labour supply functions together with the Marshallian and Hicksian functions.

The original interest in the life-cycle labour supply is motivated by the need to investigate the various dimensions of labour supply, such as the determinants of the shape of the life-cycle hours profile, the labour supply response to the aggregate wage, the changes and the source of the idiosyncratic year-to-year changes in labour supply. The existing literature, however, manages to shed little light on the original questions focusing

mainly on one aspect of the inter-temporal hours variation - the labour supply response to the wage growth along a known life-cycle trajectory, whilst ignoring other aspects. One ignored aspect is the labour supply response to wage changes under uncertainty, meaning wage changes that determine individuals to revise their expectations of their future wages (Card, 1991). MaCurdy (1983, 1985) makes the most significant contribution in incorporating uncertainty.

The life-cycle framework is proposed as an explanation for all the components of the individual labour supply (Card, 1991). Lucas and Rapping (1970) consider that the life-cycle model can be used to explain the aggregate year-to-year movements in labour supply “time effects”. Heckman (1974, 1975), Ghez and Becker (1975) claim that the life-cycle model is able to explain the systematic age effects in hours of work (“age effects”) and the differences across people with respect to their hours of work over the life-cycle “person-specific effects”. MaCurdy (1980) and Altonji (1986) formulate life-cycle models of labour supply which explain the “person-and-year specific” changes in hours of work through changes in wages.

In the context of estimating and simulating life-cycle labour supply, the choice of the labour supply elasticity to be simulated depends on the scope of the exercise. If the interest is in comparing the impact of wage variations across consumers on labour supply, the variation in the entire wage profile must be examined. Because the variation of the wage profile affects the value of the marginal utility of wealth, Frisch elasticity cannot be used to measure the effect of this variation. The estimation of the full impact on wages requires the estimation of the effect of the shifts of the wage profile on the hours of work besides the estimation of the inter-temporal elasticity.

The estimation of the parametric shift requires the specification of the impact of the wage profile on the marginal utility of wealth. The estimation of the uncompensated substitution elasticity is undertaken in a limited number of studies as it relies on specifying the functional form of the “marginal-utility-of-wealth constant” parameter, proven to be difficult in practice. Some studies (MaCurdy 1980, 1985) ignore the functional form of this parameter because of its complicated functional form in initial assets, lifetime wages, the interest rate, the rate of time preference and the “taste” parameters. They assume that the approximation of its life-cycle specification is a linear function of measured characteristics, the natural log of wages at each age, initial wealth and an unobserved random variable representing unmeasured characteristics. Others derive an expression of the marginal utility of wealth by taking into account the restrictions of the optimization process (e.g. Bover, 1989).

The estimation of the full impact of wage changes, both evolutionary and parametric, are of core importance for policy evaluation. Assuming that the tax and benefit reforms represent unanticipated shifts in net real wages today and in the future, the elasticity measuring the cumulated response to evolutionary changes and parametric shifts in the life-cycle wage profile represent the most appropriate means for describing the response to these reforms (Blundell and MaCurdy, 1981).

### **III. A Life-Cycle Model of Labour Supply**

The model in this chapter follows the theoretical specification introduced by MaCurdy (1980, 1985) and Medoff and Abraham (1981), and the unifying labour supply framework introduced by Blundell and

MaCurdy (1998). The model aims to estimate the effect of evolutionary wage changes assuming no parametric shift in the wage profiles.

The labour supply responses are estimated and simulated under two scenarios: first, assuming that individuals make labour supply decisions in a world of perfect certainty, and second, assuming that individuals make decisions in a world in which they are uncertain about their future wages.

### *1.III.1 A Life-Cycle Model of Labour Supply under Certainty*

Under the assumption of certainty, the effect of evolutionary wage changes on hours of work represents the conventional “marginal-utility-of-wealth constant” inter-temporal elasticity of labour supply obtained from the Frisch labour supply equations and the Euler condition. The marginal-utility-of-wealth parameter serves as a sufficient statistic that captures the information from the other periods needed to solve the maximization process in the current period (Blundell and MaCurdy, 1998).

The inter-temporal elasticity estimated from the Frisch specifications is relevant for predicting the individual labour supply into the future assuming a steady state, mainly due to the presence of the marginal utility of wealth, which is individualized, constant over time, and accounts for the worker’s future plans (MaCurdy, 1980).

Following MaCurdy (1980, 1983), the theory underlying the model of lifetime hours of work used in this chapter represents an extension of Friedman’s (1957) permanent income theory to a situation where the relative price of consumption and leisure varies over the life-cycle. The permanent income hypothesis can be extended to the lifetime labour supply to assume that individuals/households look into the future when deciding the current number of hours supplied on the labour market. This theory allows us to make the distinction between a consumer’s dynamic behaviour and the factors determining the differences in hours of work between consumers. This separation leads to a manageable empirical model that accommodates the differentiation between the labour supply responses to evolutionary changes and those to parametric changes in the wage profiles.

#### *1.III.1.1 An economic model of labour supply under certainty with homogenous preferences and heterogeneous individual effects*

This section presents an economic model of life-cycle labour supply decisions assuming that a worker takes his/her decisions in an environment of perfect certainty with respect to his/her future income. The worker is assumed to choose consumption and hours of work at each age to maximize a lifetime preference function, strongly separable over time, subject to wealth constraints. The model described in this chapter is designed for single decision-makers, but an extension to joint decision makers is straightforward.

In the application below, male’s labour supply behaviour is considered as independent while women’s labour supply is conditioned on other household incomes besides their own earnings. Assuming that the consumer  $i$  at the age of  $t$  has a utility given by the concave function  $U_i(C(t), L(t), X(t))$ , where  $C(t)$  is the consumption at age  $t$ ,  $L(t)$  is the number of hours of leisure at age  $t$  and  $X(t)$  is a vector of “taste shifters” variables at age  $t$ , the vector  $X(t)$  can include both observed and unobserved variables.

Due to the assumption of separating utility, the lifetime preference function can be formally presented as the sum of discounted future utilities at the moment  $t = a$ , equivalent to the beginning of the active life, where  $t$  represents age and  $a$  the age of entrance into the labour market:

$$\sum_{t=a}^T \frac{1}{(1+\rho)^t} U_t(C(t), L(t), X(t)) \quad (1)$$

The lifetime (active life) is assumed to consist of  $T - a + 1$  periods, where  $T$  represents the age of retirement. The rate of time preference used for discounting the value of future utility is represented by  $\rho$ . Formally, the consumer has to choose  $C(t)$  and  $L(t)$  at each age to maximize their lifetime preference function (1) subject to a lifetime wealth constraint:

$$A(0) + \sum_{t=a}^T R(t)H(t)W(t) = \sum_{t=a}^T R(t)C(t) \quad (2)$$

Where  $A(0)$  is the level of assets at the beginning of the active life of each consumer,  $H(t)$  the number of labour market hours at age  $t$ ,  $W(t)$  the exogenous wage rate at age  $t$ , then  $R(t) = 1 / ((1+r(1)) \dots (1+r(t)))$  is the discount rate which is used to convert the real income at age  $t$  into its equivalent for age  $a$  and  $j$  refers to the sample period. In period  $j$ , the consumer can borrow and lend at a rate of interest equal to  $r(j)$ . It is assumed that the rate of interest stays constant over time, so that the discount rate at age  $R(t) = 1 / (1+r(t))^t$ .

To create the Frisch labour supply functions, it is assumed that the contemporaneous utility function for each individual at age  $t$  takes the form:

$$U_{it} = G(C_{it}, X_{it}) - \Psi_{it}(H_{it})^\sigma \quad (3)$$

Where  $G$  is a monotonically increasing function of  $C_{it}$ ,  $\sigma$  is a time-invariant preference parameter common across consumers and  $\Psi_{it}$  is an age-specific parameter of "tastes", which depends on the consumer characteristics expected to influence his/her utility at age  $t$  (age, education, number of children, etc.). The participation decision is included in the preference parameter. The analysis assumes  $\Psi_{it}$  is related to the worker's characteristics by the function  $\Psi_{it} = \exp\{-X_{it}\psi' - v_{it}'\}$  where  $v_{it}'$  represents the contribution of unobserved characteristics and  $\psi'$  is a vector of preference parameters. Assuming an interior optimum, the implied Frisch labour supply function or marginal utility of wealth constant labour supply function is obtained from maximizing the utility in period  $t$  subject to the lifetime wealth constraint. The Lagrange function then takes the form:

$$\mathcal{L}_i = U_{it} - \lambda_i \left( A_{i0} + \sum_{t=0}^T R_t H_{it} W_{it} - \sum_{t=0}^T R_t C_{it} \right)$$

The notation is simplified by considering the age of entrance into the labour market equal to zero. The first order conditions result in the Frisch or " $\lambda$  constant" consumption and labour supply functions:

$$C_{it} = C(\lambda_i, W_{it}, X_{it}); H_{it} = H(\lambda_i, W_{it}, X_{it})$$



The first order condition with respect to hours of work at age  $t$  implies that:

$$\ln H_{it} = \alpha(\ln \lambda - \ln \sigma) + \psi X_{it} + \beta t + \alpha \ln W_{it} + v_{it} \quad (4)$$

Where  $\alpha = \frac{1}{1-\sigma}$ ;  $\psi = \alpha \psi'$ ,  $v_{it} = \alpha v'_{it}$ , and  $\beta = \alpha(\rho - r)$ . At this stage of the analysis the assumption is that  $\alpha$ ,  $\psi$  and  $\beta$  are constant across consumers and time.

The expression (4) represents the Frisch or marginal utility of wealth constant labour hours of work function. Its functional form depends on the form of the utility function and on whether a corner solution is chosen for hours of work at age  $t$  (Blundell and MaCurdy, 1999). If an individual chooses to participate in the labour market, then model (4) applies and an interior solution is assumed. If however, the decision is not to participate, the hours of work are set to zero.

The Frisch labour supply function decomposes the labour supply decision with respect to the hours of work into personal and professional characteristics ( $X_{it}$  and age) observed at time  $t$ , the wage rate at time  $t$  ( $W_{it}$ ) and the  $\lambda$  component, which represents the sufficient statistic summarizing the relevant information for each consumer from the other periods.

The optimal value of  $\lambda$  is obtained by substituting the  $\lambda$  constant consumption and labour supply functions into the budget constraint given by equation (3).  $\lambda$  is expressed as a function of initial assets, lifetime wages, interest rates, rates of time preference and tastes. In other words,  $\lambda$  summarizes the lifetime information that the consumer requires when choosing his/her optimal level of current consumption and labour supply.  $\lambda$  represents the correspondent statistic to the permanent income from Friedman's (1957) permanent income theory and represents a permanent component, which together with the current wage, determines the consumer's current consumption and labour supply.  $\lambda$  also represents a statistic capable of characterising historic and future information for lifetime wages and assets that are relevant for the current choices of consumption and labour supply. The general conclusion is that, assuming perfect certainty, the " $\lambda$  constant" consumption and labour supply functions fully characterize a consumer's dynamic behaviour (Blundell and MaCurdy, 1998; MaCurdy, 1980, 1985).

As shown by Blundell and MaCurdy (1999), in a world of perfect certainty,  $\lambda$  can be captured as an individual specific effect, which is constant over time. For this reason, for each individual, changes in wages have no impact on  $\lambda$ . In conclusion, the Frisch elasticity is the correct elasticity for assessing the impact of the evolutionary changes along the life-cycle wage profile on labour supply. The main characteristics of the  $\lambda$  component, the fact that it is specific to each individual, constant over time and incorporates the future plans of each worker in a simple way, allows the use of the  $\lambda$  constant function (4) to forecast individual labour supply outside the sample period.

Given that  $\lambda$  is time-invariant and unique for each worker, the intercept term from the labour supply equation (4) is:

$$F_i = \alpha(\ln \lambda_i - \ln \sigma)$$

This represents a time-invariant component, unique to each individual and can therefore be treated as a fixed effect in the estimation. The argument in favour of this choice stems from the content of the  $\lambda$  component

that depends on the past and future information of the variables relevant for determining the individual labour supply. Because  $\lambda$  is correlated with the exogenous variables in the model, the natural conclusion is that  $F_i$  is also correlated with the exogenous variables in the model, and therefore cannot be treated as a “random effect”. The parameters in model (4) are estimated by fixed effects. In order to circumvent the problem induced by the fixed effects specification in a simulation context, this chapter applies a three-stage procedure to decompose the unit fixed effects.

The empirical specification of the individual effects requires a formal specification of the  $\lambda$  component. The one obtained from the optimisation process is too complicated to estimate it empirically. This chapter adopts the suggestions made by MaCurdy (1981), Blundell and MaCurdy (1999). For the empirical specification,  $\ln\lambda_i$  can be approximated by:

$$\ln\lambda_i = D_i\varphi^* + \sum_{t=0}^{T_i} \gamma_t^* \ln W_{it} + A_{i0}\vartheta^* + a_i^*$$

$D_i$  is a vector of observed time-invariant characteristics or rarely changing variables,  $\varphi^*$ ,  $\gamma_t^*$ ,  $\vartheta^*$  are parameters assumed constant across consumers, and  $a_i^*$  is an error term. This empirical specification imposes strong restrictions, as it is assumed that the worker knows he/she will work  $T_i$  years, and their total lifetime income. It also incorporates the effect of interest rates and rates of time preference in the intercepts and the other parameters. Under this assumption, the time-invariant intercept of equation (4) can be approximated by:

$$F_i = \alpha \left( D_i\varphi^* + \sum_{t=0}^{T_i} \gamma_t^* \ln W_{it} + A_{i0}\vartheta^* + u_{it}^* - \ln\sigma \right) = D_i\varphi + \sum_{t=0}^{T_i} \gamma_t \ln W_{it} + a_i \quad (5)$$

$\varphi = \alpha\varphi^*$ ,  $\gamma_t = \alpha\gamma_t^*$ ,  $\vartheta = \alpha\vartheta^*$  and the intercept include the term  $(-\alpha\ln\sigma)$ , and  $A_{i0}$  represents the initial income, assumed to be zero for all individuals. The present value of lifetime income is constant for each individual and can be approximated by functions of all the personal and professional characteristics influencing a lifetime wage path as:

$$\sum_{t=0}^{T_i} \gamma_t \ln W_{it} = f(X_{i1}, X_{i2}, \dots, X_{iT_i})$$

Where  $X_{ik}$ ,  $k=1, 2, \dots, T_i$  represents a vector of personal and professional career characteristics which can be expressed as functions of two components: a time invariant part  $D_1$  and an unobserved part  $D_2$ :

$$\sum_{t=0}^{T_i} \gamma_t \ln W_{it} = f(X_{i1}(D_{i1} + D_{i2}), X_{i2}(D_{i1} + D_{i2}), \dots, X_{iT_i}(D_{i1} + D_{i2})) \quad (6)$$

Assuming that this is an additive function for  $D_1$  and  $D_2$ , the present value of the lifetime wage can be approximated by:

$$\sum_{t=0}^{T_i} \gamma_t \ln W_{it} = f(D_{i1}) + f(D_{i2})$$

By substituting equation (6) into (5), the individual effect can be presented as follows:

$$F_i = D_{i1}\varphi + f(D_{i1}) + f(D_{i2}) + a_i \quad (7)$$

The time invariant characteristics  $D$ , can be decomposed into the observed part  $D_1$  and the unobserved part  $D_2$ :

$$F_i = (D_{i1} + D_{i2})\varphi + f(D_{i1}) + f(D_{i2}) + a_i$$

The observed part can then be approximated by a linear form of  $D_{i1}$ , whereas the unobserved part, assumed to follow a normal distribution, is incorporated in the error term as:

$$F_i = D_{i1}\varphi + f(D_{i1}) + D_{i2}\varphi + f(D_{i2}) + a_i \quad (8)$$

$$F_i = D_{i1}\varphi^* + c_i \quad (9)$$

Therefore, the individual unit effect can be approximated by a linear function of the time-invariant characteristics plus a normally distributed error that accounts for the unexplained part of the individual effect. Substituting equation (9) into (4), the hours of work process can be approximated as follows:

$$\ln H_{it} = D_{i1}\varphi^* + \psi X_{it} + \beta t + \alpha \ln W_{it} + c_i + v_{it} \quad (10)$$

The simulation of labour supply based on the economic model given in (4) requires also a mechanism for predicting lifetime wage profiles. It is assumed that the lifetime wage profile is:

$$\ln W_{it} = L_i + \pi_{0,k} P_{it} + \pi_1 t + \pi_2 t^2 + \varepsilon_{it} \quad (11)$$

$L_i$  is the individual specific effect,  $P_{it}$  is a vector of personal and professional career characteristics,  $\pi_{0,k}$  is a vector of coefficients which are constant across time and individuals,  $k$  is the number of personal and professional career characteristics in the model and  $\varepsilon_{it}$  the error term. Experience and education are assumed to be exogenously determined.

### 1.III.1.2 An expansion of the economic labour supply model under certainty with heterogeneous preferences

In this section, the model is extended by allowing for increased heterogeneity amongst the individuals, who are assumed to have different preferences on their labour supply. The preference parameter  $\sigma \sim N(0, \Gamma)$  is individual specific but time invariant.  $\alpha_i, \beta_i, \alpha(\ln \lambda_i - \ln \sigma_i)$  are assumed to be individual specific and follow normal distributions across individuals. Given these assumptions, the lifetime labour supply function is:

$$\begin{aligned} \ln H_{it} &= \alpha_i (\ln \lambda_i - \ln \sigma_i) + \psi X_{it} + \beta_i t + \alpha_i \ln W_{it} + v_{it}, \\ \ln H_{it} &= (\zeta_0 + \varsigma_{0,i}) + \psi X_{it} + (\zeta_1 + \varsigma_{1,i}) t + (\zeta_2 + \varsigma_{2,i}) \ln W_{it} + v_{it} \end{aligned} \quad (12)$$

$\alpha(\ln \lambda_i - \ln \sigma_i) = \zeta_0 + \varsigma_{0,i}$  is the individual effect constant over time,  $\beta_i = \zeta_0 + \varsigma_{1,i}$  the random coefficient of age and  $\alpha_i = \zeta_2 + \varsigma_{2,i}$  the random coefficient of wage.  $\zeta$  is the mean intercept of hours of work,  $\beta$  the mean slope of age,  $\alpha$  the mean slope of the natural logarithm of wage and  $\varsigma_{k,i}$ ,  $k \in [0, 2]$  the deviation of individual intercept

and slopes from the mean values of  $\zeta_0, \zeta_1, \zeta_2$ .  $\psi$  is a vector of fixed coefficients, constant across individuals and time.

In addition, in order to allow for increased heterogeneity also in wage, the lifetime wage path has the following specification:

$$\ln W_{it} = (\pi_0 + \xi_{0,i}) + R_{it}(\pi_1 + \xi_{1,i}) + (\pi_2 + \xi_{2,i})t + \pi_3 P_{it} + \pi_4 t^2 + \varepsilon_{it} \quad (13)$$

$R_{it}$  and age represents the variables with random coefficients,  $\pi_0$  represents the mean intercept of wage,  $\pi_k, k=1,2$  is a vector of coefficients illustrating the mean slopes of  $R_{it}$  and age, and  $\xi_{k,i}, k=0,1,2$  represents the deviation of the individual intercepts and slopes from the mean values of  $\pi_0, \pi_1$  and  $\pi_2$ .

### 1.III.2 An expansion of the economic labour supply model under uncertainty

The above economic models of life-cycle consumption and labour supply behaviours are assumed to be certain at the time period 0, meaning that agents have perfect information and perfect predictions in the beginning. The following model extension relaxes the assumption of certainty as individuals are assumed to act in a world of uncertainty, where they adjust their expectations every period based on current and past information. This uncertainty is accounted for in the model by including a forecasting error which incorporates the past and future differences between realized values and their expectations.

It is assumed that agents form an expectation of wage adaptively based on what has happened in the past. Due to the imperfection of forecasting, the wealth constraint needs to be updated every time period. Therefore the lifetime wealth constraint at time period  $k$  in time period 0 with a value of  $WC_0(k)$ , can be expressed as the sum between the assets in period 0, the discounted sum of the total earnings between period 0 and period  $k$ , and the sum of the expected total earnings, with discount factor  $d(t)$  between period  $k$  until retirement (period  $T$ ):

$$WC_0(k) = A(0) + \sum_{t=0}^k d(t)H(t)W(t) + \sum_{t=k+1}^T d(t)E(W(t)H(t)) \quad (14)$$

Alternatively, it can be expressed as the sum between the assets in period 0, the expected discounted total earnings over the active life formed before entering the labour market and an adjustment error. The adjustment error equals the sum between the difference between the realised earnings and their expectations made before entering the labour market over the period 0 to  $k$  (part A), and the expectations adjustment of the total earnings made in the period  $t+1$  to  $T$  (Part B).

$$\begin{aligned} WC_0(k) &= A(0) + \sum_{t=0}^T d(t)(E(H(t))|I_{-1})(E(W(t))|I_{-1}) \\ &\quad + \underbrace{\sum_{t=0}^k (d(t)H(t)W(t) - d(t)(E(H(t)W(t))|I_{-1}))}_A \\ &\quad + \underbrace{\sum_{t=k+1}^T (d(t)(E(H(t)W(t))|I_k) - d(t)(E(H(t)W(t))|I_{-1}))}_B \\ &= WC_0(0) + \mathfrak{U}_k \end{aligned} \quad (15)$$

$I_t$  represents all the information available to an agent at time period  $t$ , which includes personal characteristics and labour supply activities, and  $\mathfrak{U}_k$  represents the error adjustment term. This suggests that current expectations of life time wealth reflects past expectations, and an error adjustment term that could either lower or raise the total constraint. To further simplify the equation, it is assumed that the interest rate is certain. Therefore, the error adjustment term  $\mathfrak{U}$  is expressed as:

$$\mathfrak{U}_t = f(v_t(W), v_t(H), v_t(X)) = f(v_{t-1}(W), v_{t-1}(H), v_t(X), W_t, H_t(W_t)) \quad (16)$$

$v_t(W)$  is a vector of all wage information until the time period  $t$ ,  $v_t(H)$  is a vector of working hour information until time period  $t$ , and  $v_t(X)$  is a vector of personal characteristics information until time period  $t$ . As the working hour is modelled using the wage rate and the personal characteristics, the new expectation adjusted lifetime wealth constraint can be approximated as:

$$WC_0(k) = WC_0(0) + \mathfrak{U}_k \approx WC_0(0) \left( p e^{\sum_{t=0}^k (d(t)(W(t) - (E(W(t))|I_{k-1}))) + \hbar X_{it}} \right) \quad (17)$$

$p$  is an approximation parameter,  $X_{it}$  is a vector of personal characteristics which affect the form of the current expectation, and  $\hbar$  is the vector of coefficients of personal characteristics.  $\ln(WC_0(k)) - \ln(WC_0(0))$  can be formulated as:

$$\ln(WC_0(k)) - \ln(WC_0(0)) = \mu \hat{A}_k + \hbar X_{it}$$

$$\hat{A}_k = \sum_{t=0}^k \left( \frac{1}{\prod_{j=0}^k (1+r(j))} (W(t) - (E(W(t))|I_{k-1})) \right)$$

Combining the anticipated paths for wages and income with the approximated empirical Lagrange function modifies the lifetime constraint. Therefore, the  $\lambda$  component takes the form:

$$\ln \lambda_{it} = D_i \varphi^* + \sum_{t=0}^{T_i} \gamma_t^* \ln W_{it} + \mu \hat{A}_k + \hbar X_{it} + A_{i0} \mathfrak{S}^* + a_i^* \quad (18)$$

$\mu$  is an adjustment ratio, which shows the marginal change of working hours caused by the error term  $\left( \frac{\partial H_t}{\partial \hat{A}_k} \right)$ .  $\hbar$  explains the role of personal characteristics in the formulation of expectation. It is not possible, however, to distinguish  $\hbar$  from the personal character coefficient vector  $\psi$  in equation (12).

#### IV. Empirical Implementation

This section introduces the econometric and the simulation techniques for estimating and simulating lifetime labour supply, both under certainty and uncertainty. The model is structured in three parts: a selection model, a model for lifetime wage profiles and a model for lifetime labour supply profiles.

### 1.IV.1 Data

This study uses the German Socio Economic Panel (GSOEP) data from 1996 to 2003 for estimation and 2004-2005 data for the simulation validation exercise. Only adults respondent who have not reached the retirement age are included in the study. Some basic employment variable information is presented in Table 1.

**Table 1 Basic Descriptive of Some Employment Variables**

<i>Variable</i>	<i>Mean</i>	<i>Std. Dev.</i>
Age	40.52	12.78
Number of children	0.94	1.00
Employed in formal sector	0.50	0.50
Fulltime experience	1.63	1.49
Part-time experience	0.24	0.64
Unemployment period	0.39	0.79
Health situation	0.08	0.28
Hours of working	19.78	20.78
Education (Years)	11.43	2.95
Total number of observations	11456	

The dependent variable is the logarithmic value of actual working hours per week. The wage rate in the estimation and the simulation is calculated using the actual working hours. “Participation” is defined as employment in the formal labour market, except for vocational training, zero working hours, military service and community service, which are all modelled as “non-participation” in the selection model.

### 1.IV.2 The Empirical Model of Lifetime Labour Supply under Certainty

This subsection develops the econometric model for the labour supply under certainty. The empirical model follows the structure of the economic model: in the first stage it assumes a homogenous preference parameter  $\sigma$ , and in the second it allows for a higher degree of heterogeneity amongst individuals.

#### 1.IV.2.1 An Econometric Model of Labour Supply under Certainty with Homogenous Preferences

For estimating a lifetime labour supply which incorporates a lower degree of heterogeneity this chapter proposes a model similar to Plumper and Troeger (2007) and Hsiao (2003). The model extends the standard fixed effects specification to an estimation procedure which enables an efficient estimation of time-invariant and rarely changing variables by applying a “fixed effects vector decomposition”.

The lifetime labour supply model follows the structure of the economic model in (4):

$$y_{it} = \zeta + \sum_{k=1}^K \beta_k X_{kit} + \sum_{c=1}^C \rho_c D_{ci} + u_i + v_{it} \quad (19)$$

$y_{it}$  is the natural logarithm of hours worked per week,  $\zeta$  is the intercept of the base unit,  $X$  are the time-varying variables (the natural logarithm of wage, age, age squared, children, health dummies, household type dummies, cumulated experience until last year (full time, part-time, unemployment), other household

income (for women only), sector dummies),  $D$  are the time-invariant variables (education, education squared, cohort dummies),  $u_i$  are the unobserved individual specific effects,  $v_{it}$  is an i.i.d. error term, and  $J$  and  $\varrho$  are parameters common to all individuals and constant over time.

The lifetime labour supply model is estimated using a three-stage procedure. To start, the model (19) is estimated using the standard fixed effects estimator and it is possible to extract the individual fixed effects or unit effects as:

$$\hat{u}_i = \bar{y}_i - \sum_{k=1}^K \bar{\varrho}_k^{FE} \bar{x}_{ki} - \bar{v}_i \quad (20)$$

The estimated individual specific effect  $\hat{u}_i$  captures the unobserved individual-specific effect, the observed individual-specific effects  $D$ , the individual means of the residuals  $\bar{v}_i$  and the individual means of the time-varying variables. In the second step, the estimated individual specific effect is regressed against the observed time-invariant characteristics and the rarely changing variables to obtain the unexplained part of the individual-specific effects. The individual specific effects are decomposed as follows:

$$\hat{u}_i = \sum_{c=1}^C \varrho_c D_{ci} + h_i \quad (21)$$

$\sum_{c=1}^C \varrho_c D_{ci}$  is the explained part.  $h_i$  is the residual from equation (21) and captures the unexplained part of the individual-specific effect:

$$h_i = \hat{u}_i - \sum_{c=1}^C \varrho_c^{OLS} D_{ci} \quad (22)$$

In the third stage, the individual effect from the model (19) is substituted with the unexplained part of the decomposed individual fixed effect vector obtained in the previous stage, resulting in an error that is no longer correlated with the time varying covariates included in the model. Therefore, the model (23) can be estimated consistently by pooled OLS:

$$y_{it} = \zeta + \sum_{k=1}^K \varrho_k X_{kit} + \sum_{c=1}^C \varrho_c D_{ci} + \rho h_i + v_{it} \quad (23)$$

The Monte Carlo simulations conducted by Plumper and Troeger (2007) reveal the circumstances under which the FEVD is inferior to the pooled OLS, random effects (RE) and fixed effects (FE). OLS is more appropriate when there are no individual effects, RE when the individual effects are uncorrelated with the other explanatory variables, and FE when the assumptions of the RE are violated and the within-variance of the variables of interest is sufficiently large compared with the between-variance. If this condition does not occur, the vector decomposition technique has better finite sample properties in estimating models that have time-invariant or rarely changing variables correlated with individual-specific effects.

The lifetime wage process is estimated based on model (24) using the same procedure:

$$w_{it} = \pi_0 + X_{it}' \pi_1 + D_i' \pi_2 + \xi_{0,i} + \varepsilon_{it} \quad (24)$$

$w_{it}$  is the natural logarithm of gross hourly wage is,  $X_{it}$  is a vector of time-varying variables, and  $D_i$  a vector of time invariant characteristics.  $\xi_{0,i}$  is the individual-specific effect and  $\varepsilon_{it}$  the error term. The time-varying covariates in the wage equations are age, age squared, children, health dummies, household type dummies, cumulated unemployment experience until last year, and sector dummies. The time-invariant variables are education, education squared and cohort dummies.

When analysing the wage and the labour supply processes using panel data, the natural question that arises is whether the non-response or the missing observations are endogenously determined.

One source of sample selectivity is the unbalanced nature of the panel. If the panel attrition is endogenous to the wage and the hours processes, sample selection would be informative for wage and hours, and therefore the estimates from the wage and hours equations would be biased. The panel attrition bias is disregarded from the analysis because previous studies using the GSOEP have indicated that the sample selection bias is not significant (Galler, 1996; Rendel and Buechel, 1994).

Another sample selectivity bias comes from the fact that wages and working hours are observed only for the individuals in the labour market and selection bias is determined by the differences between workers and non-workers. If the sample under analysis is randomly selected, it is assumed that both workers and non-workers have similar observed and unobserved characteristics and the selection process does not bias the estimates obtained using the working sample. In contrast, if the decision to work is no longer random and people select themselves into the labour market based on certain characteristics, then it is reasonable to assume that workers and non-workers have different observed and unobserved characteristics. A selection bias arises when some component of the participation decision is relevant to the wage and hours processes. Disregarding these relationships, the estimates of wages and hours of work from the subsample of working individuals will be biased.

If the relationships between the participation decision and the wage and hours processes occur through the observables, the selection bias can be controlled by introducing the appropriate conditioning variables in the wage and hours equations. If the relationships between the participation decision and the wage and hours processes occur through the unobservable, meaning that the unobserved characteristics affecting the participation decision are correlated with the unobservable from the wages and hours equations, simply controlling for the observables is not enough to obtain unbiased estimates. If the observables are correlated with the unobservable, in order to get unbiased estimates, the wage and hours equations should include an estimate for the unobservable (Vella, 1998).

The selection model is defined using the latent variable model:

$$s_{it}^* = X_{it}'\vartheta + D_i'\varrho + u_i + v_{it} \quad (25)$$

The selection indicator  $s_{it}$  defines the observed employment status  $y_{it}$ :

$$y_{it} = \begin{cases} 1 & \text{if } s_{it}^* > 0 \\ 0 & \text{if } s_{it}^* \leq 0 \end{cases} \quad (26)$$

For fixed  $u_i$  the probability of observing  $y_{it} = 1$  is given by:



$$P\{y_{it}=1 | X_{it}, D_i\} = \Phi(u_i + X_{it}'\vartheta + D_i'\varrho) \quad (27)$$

$\Phi(\cdot)$  is the cumulative distribution function and  $v_{it}$  is assumed to be normally distributed. With i.i.d. error terms, the log likelihood function for a fixed effect probit model is given by:

$$\begin{aligned} \log L(\vartheta, \varrho, u_1, u_2, \dots, u_N) &= \sum_{i,t} y_{it} \log \Phi(u_i + X_{it}'\vartheta + D_i'\varrho) \\ &\quad + \sum_{i,t} (1 - y_{it}) \log \Phi(u_i + X_{it}'J + D_i'\varrho) \end{aligned}$$

For a fixed  $T$  and  $N \rightarrow \infty$ , the maximum likelihood is inconsistent because the number of unknown parameters grows with the number of individuals in the sample. This is the so called “incidental parameter problem” which impedes the fixed effects probit model to be estimated consistently for a fixed number of periods. To circumvent this problem and to estimate the selection model accounting for the unobserved heterogeneity, the model decomposes the unobserved effect  $u_i$ , following the approach introduced by Mundlak (1978). The assumption is that the unobserved effect can be modelled as:

$$u_i = \eta + \overline{X}_i' \tau + a_i \quad (28)$$

This equation assumes that the correlation between  $u_i$  and  $x_{it}$  acts only through the time averages of the exogenous variables, whereas  $a_i$  represents the remaining part of the unobserved effect which is independent of the time-varying variables. Equation (28) can be substituted into the selection model as follows:

$$y_{it} = \begin{cases} 1 & \text{if } \eta + X_{it}'\vartheta + Z_i'\varrho + \overline{X}_i'\tau + a_i + v_{it} > 0 \\ 0 & \text{if } \eta + X_{it}'\vartheta + Z_i'\varrho + \overline{X}_i'\tau + a_i + v_{it} \leq 0 \end{cases} \quad (29)$$

$$v_{it} | X_i, Z_i \sim N(0,1), t=1, \dots, T \quad (30)$$

To summarize, the selection model includes time-invariant characteristics (education, education squared, cohort dummies, nationality dummies), time-varying variables (age, age squared, interaction between education and age, cumulated experience (full time, part-time, unemployment), health dummies, household type dummies, children and other household income (for women only)) and the means for the time-invariant variables (except for age).

To test for selection bias in both hours and wage equations for men and women, the study applied the approach introduced by Wooldridge (1995) for the FE specification. A standard probit was run for each time period based on the model (30). For people participating in the labour market, the inverse Mills ratios were computed and then introduced into the initial models for hours and wages. The final step involved estimating the augmented models by applying the FE specification and testing the coefficient of the inverse Mills ratios.

To correct for the sample selection in the wage and hour equation, the inverse Mills ratios for each year were included as time varying variables in the FEVD estimation of the wage and hours models.

For simulating both labour supply and wage processes, the main requirement imposed on the estimation method is to provide consistent and unbiased estimates. Allowing for unobserved heterogeneity and for a

correlation between the unobserved individual effects and the other explanatory variables requires the use of FE estimation, however this estimation technique is rather useless in a simulation context. A good alternative is the FEVD technique, which circumvents the problems of a standard FE model (i.e. assumes a correlation structure between the unobserved individual effects and the other explanatory variables) by decomposing the individual effect and estimating the last stage as a pooled OLS.

#### 1.IV.2.2 An Econometric Model of Labour Supply under Certainty with Heterogeneous Preferences

For the estimation of the lifetime labour supply model which incorporates a high degree of individual heterogeneity, the study proposes a mixed fixed and random coefficient model similar to Rabe-Hesketh and Skrondal (2005). The fixed coefficients are considered constant across consumers and time, and the random coefficients are unique for each individual, but constant over time.

The assumption is that the labour supply model can be approximated by:

$$y_{it} = \zeta_0 + J_{it}'\boldsymbol{\theta} + (\zeta_{0,i} + v_{it}), i = 1, \dots, N; t = 1, \dots, T \quad (31)$$

$J_{it}$  is a  $1 \times (K + M)$  vector of explanatory variables,  $\boldsymbol{\theta}$  is the corresponding vector of coefficients,  $\zeta_{0,i}$  is the individual-specific effect, and  $v_{it}$  the error term. As the  $\zeta_{0,i}$  component induces a within-individual dependence, the composite error term  $(\zeta_{0,i} + v_{it})$  is correlated over time. Serially correlated errors imply then the OLS or maximum likelihood standard errors are no longer valid. This dependence can be taken into account either by using the sandwich estimator for the standard errors, which does not make any assumptions about the distribution of within-dependence of the residuals, or by modelling the dependence explicitly (Rabe-Hesketh and Skrondal, 2005).

One way to model dependence is to decompose the error component into a time-constant or a permanent error component  $\zeta_{0,i}$ , unique for each individual, and a transitory error component  $v_{it}$ , which varies across individuals and time. The permanent error component represents the combined effect of the omitted time-constant covariates and of the unobserved heterogeneity.  $\zeta_{0,i}$  and  $v_{it}$  are assumed normally distributed, independent of each other.  $v_{it}$  is assumed independent across individuals and over time. Model (31) can be rewritten by moving  $\zeta_{0,i}$  in the intercept:

$$y_{it} = (\zeta_0 + \zeta_{0,i}) + J_{it}'\boldsymbol{\theta} + v_{it} \quad (32)$$

Model (32) represents a random-intercept model, which is a regression model with an individual specific intercept.  $\zeta_{0,i}$  can be interpreted as a random parameter that is estimated together with the variance of the  $v_{it}$ . The parameters of the random-intercept model are estimated by maximum likelihood. To conclude, the linear random-intercept model allows the overall level of hours to vary across individuals after controlling for covariates.

Additional heterogeneity is incorporated by including additional random coefficients besides the random intercept, meaning that the effects of some covariates are allowed to vary across individuals. The resulting model is a mixed fixed and random coefficient model. By introducing individual-specific slopes, the assumption of parallel individual-specific regression lines is relaxed and our model becomes:

$$y_{it} = (\zeta_0 + \zeta_{0,i}) + X'_{it}\psi + R'_{it}(\zeta_k + \zeta_{k,i}) + v_{it}, k = 1, 2 \quad (33)$$

$\zeta_0$  represents the mean intercept of hours of work,  $\zeta_k, k=1,2$  the mean slopes of the covariates chosen to have random coefficients, and  $\zeta_{k,i}, k \in [0,2]$  the deviation of individual intercept and slopes from the mean values of  $\zeta_k$ .  $\psi$  is a vector of fixed coefficients, constant across individuals and time  $X'_{it}$  represents a  $1 \times K$  vector of covariates with fixed coefficients, whereas  $R'_{it}$  is a  $1 \times M$  vector of covariates with random coefficients.

The wage model follows a similar specification; besides the random individual-specific intercept, the model is specified with two additional random coefficients for age and education:

$$w_{it} = (\pi_0 + \xi_{0,i}) + R'_{it}(\pi_1 + \xi_{1,i}) + P'_{it}\pi_2 + \varepsilon_{it} \quad (34)$$

$R_{it}$  is the vector of covariates with random coefficients.  $P_{it}$  is the vector of covariates with fixed coefficients.  $\pi_0$  is the mean intercept of wage,  $\pi_1$  the vector of coefficients illustrating the mean slopes of the  $R_{it}$  variables and  $\xi_{k,i}, k=0,1$  represents the deviation of individual intercept and slopes from the mean values of  $\pi_0, \pi_1$ .

Given the high computation costs of estimating the selection model using a random coefficient specification, only fixed coefficient specification is used in the study. The inverse Mills ratios from the estimated probit regressions are included in the random coefficients wage equation of both men and women, whereas for hours, the additional term is included only for women.

#### 1.IV.3 The Empirical Model of Lifetime Labour Supply under Uncertainty

The extension of the empirical model of lifetime labour supply to incorporate uncertainty is straightforward. The methodology is the same as the one under certainty, except for an additional regressor that captures the forecasting error. This term is introduced to incorporate that individuals adapt their expectations regarding their future wages each period.

##### 1.IV.3.1 An Econometric Model of Labour Supply under Uncertainty with Homogenous Preferences

Under uncertainty, the labour supply model assuming homogenous preferences is expressed as:

$$y_{it} = \zeta + X'_{it}\varrho + D'_i\varrho + \mu \ln \bar{\omega} + u_i + v_{it} \quad (35)$$

$$\ln \bar{\omega} = \sum_{k=a}^t (\ln w_{ik} - \ln \widehat{w}_{it} \widehat{p}_{ik}) \frac{1}{\prod_{j=a}^t (1+r(j))} \quad (36)$$

$\ln \bar{\omega}$  represents the forecasting error equal to the cumulated discounted difference between the actual wage and the expected wage multiplied by the probability of being employed, from the start of the active life ( $a$ ) until the current year. The selection and the wage models are maintained as under certainty, assuming homogenous preferences.

##### 1.IV.3.2 An Econometric Model of Labour Supply under Uncertainty with Heterogeneous Preferences

Under uncertainty, the labour supply model assuming heterogeneous preferences is expressed as under certainty, with the assumption that the forecasting error in equation (36) has a random coefficient. This

change implies that the effects of the variables vary across individuals, but the selection and wage models remain the same as under certainty. Table 2 summaries the differences of all model variants proposed.

**Table 2 An Overview of All Models Described**

<i>Models</i>	<i>Estimation Method for Labour Supply</i>	<i>Uncertainty in Future Wages</i>
Standard Random Effect (Heckman extended) Model	Random Effects Model	No
Proposed Labour supply model with homogenous preferences	FEVD	No
Proposed Labour supply model with heterogeneous preferences	Mixed Coefficients Model	No
Proposed Labour supply model with homogenous preferences with uncertain extension	FEVD	Yes
Proposed Labour supply model with heterogeneous preferences with uncertain extension	Mixed Coefficients Model	Yes

*All variables included in the estimations are the same except when the adjustment variable under uncertainty cases*

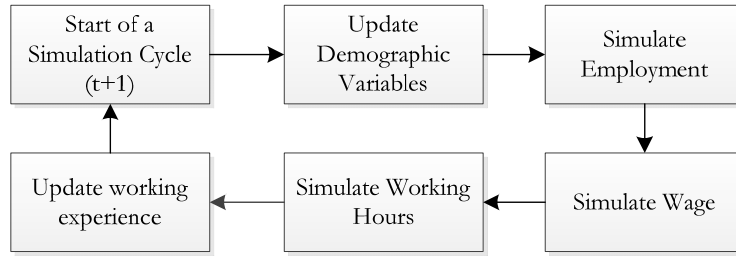
#### 1.IV.4 Simulations

This section describes how the models developed are applied to simulate the labour supply responses for 2004 and 2005. The accuracies of the projections obtained from these models are compared with the simple extended Heckman model that is commonly used in simulating continuous labour supply. Since the GSOEP dataset is a panel, the Heckman model is extended to incorporate the unobserved heterogeneity. While the estimation of the probit model is identical with one used by the selection model, the second step of the Heckman procedure is estimated by a standard RE model.

The simulations are performed using the estimates from the empirical models presented in the previous sections. The simulation follows the basic structure of a dynamic microsimulation. To simplify the exercise, this simulation only consists of the demographic and labour market module. The demographic module updates some basic demographic variables like age over time. Besides, it also updates the variables that may interact with the demographic variables. The labour market module is the core part of the simulation, which updates employment status, wage, and the hours of work. Figure 1 illustrates the steps of simulation.

For the employment model (selection model), the simulation uses pooled probit as yearly probit is unfeasible in a simulation exercise. This also circumvents the problem that the inverse Mills ratios cannot be updated during simulation. The number of formally employed individuals is aligned with the real data in 2004 and 2005. The wage and the hours simulations are based on the updated personal characteristics and are the result of the selection simulation including the selection correction. The results of the wage simulation are used in the working hour simulation.

**Figure 1 An Overview of Simulation Steps**



The simulation follows the sequence as follows: It starts by determining the value of demographic variables in the new time period and updates the related variables. For the personal characteristics that are not influenced by employment, the simulation uses the actual characteristics observed in 2004 and 2005. The simulation then moves on to the next step where labour market variables are simulated. It predicts the probability of employed given the employment selection model. Afterwards, wage can be predicted using previously estimated equations. With the wage and personal characteristics information, it is now possible to simulate the hours of work. In the case of models with uncertainty extension, correction terms were calculated right after the predicted wage becomes available. Lastly, the variables that reflect the labour trajectory are updated, this include working experience (full time, part-time, unemployed) etc. The projections of the models are compared with the actual observed hours of work for 2004 and 2005.

## V. Results

This section presents the estimation and simulation results. All models, including wage and labour supply models are estimated separately for men and women. Simulation is conducted under a simple framework as described in earlier section.

### 1.V.1 Estimation Results

The parameter estimates for the probit models are largely skipped as the estimates are in line with previous findings and all key variables are significant. Inverse Mill's ratios were included in wage and labour supply model with the exception of male labour supply equation, where the inverse Mill's ratio is not significant using any estimation method. For the random coefficients models, a test was performed to verify whether the random intercept model is sufficient to capture the heterogeneity in the wage and hours estimation. The likelihood-ratio test suggests that the random coefficients model fits better than the simple random intercept model, both for wage and labour supply.

Table 3 and 5.4 show the parameter estimates for the wage equations for women and men. Both model specifications are standard. The age effect is as expected for both men and women in all three models: a positive impact with decreasing marginal effects, showing that hourly wage has the standard humped-shaped age pattern. When estimated using FEVD, the wage profile of women shows a stronger curvature than for men. The larger coefficients of the linear and quadratic age variables show that the growth rate of wages is higher for women at younger ages, but the growth rate reduces more rapidly at later ages than for men. After the introduction of heterogeneity, the impact of age is reduced for women, whereas for men it is increased. At younger ages, however, both men and women have a similar curvature of the wage profile. The estimated rate of return on education differs quite a lot between men and women and between the three models. The model

estimated with FEVD illustrates a positive return for education for women, whereas for the other models the return appears not to be significant. For men, the returns appear to be negative in the FEVD variant model and insignificant in the other models.

Unemployment experience has a significant negative impact on wages for both men and women. The model with FEVD method illustrates the highest absolute impact, while the random coefficient model and the RE model show similar effects. In all three models, having or not having children influences wages negatively, and the impact in absolute value is higher for women than for men. The estimates of the inverse Mills ratios imply small positive correlations between the individual-specific error components of the selection model and the wage equation for men in all three models. For women, these correlations are negative in the mixed and the extended Heckman model.

**Table 3 Regression Estimates for Wage Function of Women 1996 - 2003**

<i>Variable</i>	<i>Fixed Effect Vector Decomposition</i>		<i>Mixed Fixed and Random Coefficients</i>		<i>Random-effects GLS (Extended Heckman)</i>	
	<i>Coefficient</i>	<i>Std. Error</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>Coefficient</i>	<i>Std. Error</i>
Age	0.1668	0.0061	0.1389	0.0109	0.1548	0.0073
Sd (Age)			0.0193	0.0010		
Age Squared	-0.0016	0.0001	-0.0015	0.0001	-0.0015	0.0001
Unemployment Experience	-0.0942	0.0024	-0.0772	0.0058	-0.0764	0.0057
Education-age interaction	-0.0004	0.0001	0.0006	0.0003	0.0007	0.0003
Children	-0.0721	0.0062	-0.0748	0.0104	-0.0698	0.0104
Control for Health	Yes		Yes		Yes	
Control for Sector	Yes		Yes		Yes	
Control for Nationality	Yes		Yes		Yes	
Control for Year Effect	Yes		Yes		Yes	
Inverse Mills Ratio	0.0694	0.0089	-0.0069	0.0131	-0.0126	0.0131
Education	0.0237	0.0104	-0.0337	0.0337	-0.0294	0.0311
Sd (Education)			0.0334	0.0094		
Education Squared	0.0022	0.0004	0.0028	0.0012	0.0024	0.0011
Control for Cohort Effect	Yes					
Constant			1.4544	0.3782		
Sd (Constant)			1.1352	0.0789		
Number of observations			14251		14251	
Number of groups			2767		2767	
R <sup>2</sup> (overall)					0.2969	

**Table 4 Regression Estimates for Wage Function of Men 1996 – 2003**

<i>Variable</i>	<i>Fixed Effect Vector Decomposition</i>		<i>Mixed Fixed and Random Coefficients</i>		<i>Random-effects GLS (Extended Heckman)</i>	
	<i>Coefficient</i>	<i>Std. Error</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>Coefficient</i>	<i>Std. Error</i>
Age	0.1177	0.0044	0.1393	0.0082	0.1563	0.0059
Sd (Age)			0.0202	0.0009		
Age Squared	-0.0015	0.0000	-0.0015	0.0001	-0.0014	0.0001
Unemployment Experience	-0.0973	0.0022	-0.0705	0.0056	-0.0694	0.0055
Education-age interaction	0.0023	0.0001	0.0013	0.0003	0.0012	0.0002
Children	-0.0304	0.0046	-0.0154	0.0078	-0.0159	0.0078
Control for Health	Yes		Yes		Yes	
Control for Sector	Yes		Yes		Yes	

Control for Nationality	Yes		Yes		Yes	
Control for Year Effect	Yes		Yes		Yes	
Inverse Mills Ratio	0.0503	0.0126	0.0191	0.0168	0.0078	0.0167
Education	-0.1228	0.0084	-0.0831	0.0306	-0.0728	0.0288
Sd (Education)			0.0242	0.0120		
Education Squared	0.0026	0.0003	0.0029	0.0011	0.0026	0.0011
Control for Cohort Effect	Yes		Yes		Yes	
Constant			1.7804	0.2929		
Sd(Constant)			0.8657	0.0735		
Number of observations	17484		17484		17484	
Number of groups	2999		2999		2999	
R <sup>2</sup> (overall)	0.7793				0.3753	

**Table 5 Regression Estimates for Hour Function of Women 1996 - 2003**

Variable	Fixed Effect Vector Decomposition				Mixed Fixed and Random Coefficients				Extended Heckman	
	Certainty		Uncertain		Certainty		Uncertain		Certainty	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Wage	-0.2765	0.0051	-0.2742	0.0063	-0.1989	0.0151	-0.1998	0.0158	-0.2396	0.0082
(Random Coeff sd)					0.5259	0.0131	0.5286	0.0136		
Age	0.0367	0.0057	0.0267	0.0058	0.0019	0.0159	0.0061	0.0156	0.1174	0.0077
(Random Coeff sd)					0.0343	0.0011	0.0339	0.0011		
Age Squared	-0.0008	0.0001	-0.0007	0.0001	-0.0005	0.0001	-0.0005	0.0001	-0.0005	0.0001
Experience Full time	0.0164	0.0005	0.0184	0.0005	0.0365	0.0021	0.0365	0.0021	0.0305	0.0017
Experience Part-time	0.0359	0.0008	0.0327	0.0007	0.0239	0.0027	0.0200	0.0028	0.0141	0.0022
Experience unemployment	-0.1101	0.0024	-0.1269	0.0024	-0.0089	0.0069	-0.0145	0.0073	-0.0280	0.0061
Education age interaction	0.0012	0.0001	0.0008	0.0001	0.0007	0.0004	0.0003	0.0005	0.0009	0.0003
Have child	-0.0682	0.0156	-0.0850	0.0162	-0.0512	0.0200	-0.0665	0.0212	-0.0710	0.0230
Other household income	-0.0126	0.0010	-0.0124	0.0011	-0.0089	0.0011	-0.0080	0.0011	-0.0142	0.0014
Education level	-0.0050	0.0097	-0.0062	0.0100	-0.0488	0.0418	-0.0594	0.0431	-0.0755	0.0339
Education Squared	0.0003	0.0003	0.0007	0.0004	0.0029	0.0015	0.0036	0.0015	0.0030	0.0012
Inverse Mills Ratio	0.0684	0.0099	0.0068	0.0095	-0.0071	0.0122	-0.0218	0.0129	-0.0219	0.0140
Wage Expectation Correction			-0.0001	0.0000			0.0000	0.0000		
(Random Coeff sd)							0.0008	0.0001		
Intercept (Random sd)					2.5242	0.0678	2.5155	0.0732		
Control for Health	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Year effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Household Type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Cohort Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

**Table 6 Regression Estimates for Hour Function of Men 1996 - 2003**

<i>Variable</i>	<i>Fixed Effect Vector Decomposition</i>				<i>Mixed Fixed and Random Coefficients</i>				<i>Extended Heckman</i>	
	<i>Certainty</i>		<i>Uncertain</i>		<i>Certainty</i>		<i>Uncertain</i>		<i>Certainty</i>	
	<i>Coeff.</i>	<i>Std. Error</i>	<i>Coeff.</i>	<i>Std. Error</i>	<i>Coeff.</i>	<i>Std. Error</i>	<i>Coeff.</i>	<i>Std. Error</i>	<i>Coeff.</i>	<i>Std. Error</i>
<b>Wage</b>	-0.3351	0.0033	-0.3330	0.0035	-0.2099	0.0090	-0.2218	0.0092	-0.2796	0.0052
<b>(Random Coeff sd)</b>					0.3471	0.0071	0.3435	0.0073		
<b>Age</b>	0.1235	0.0031	0.1200	0.0031	0.0520	0.0053	0.0512	0.0054	0.1583	0.0040
<b>(Random Coeff sd)</b>					0.0173	0.0005	0.0178	0.0005		
<b>Age Squared</b>	-0.0008	0.0000	-0.0007	0.0000	-0.0006	0.0000	-0.0006	0.0000	-0.0008	0.0000
<b>Experience Full time</b>	-0.0633	0.0007	-0.0634	0.0007	-0.0028	0.0015	-0.0018	0.0016	-0.0033	0.0017
<b>Experience Part-time</b>	-0.0726	0.0012	-0.0799	0.0013	-0.0296	0.0030	-0.0278	0.0032	-0.0501	0.0035
<b>Experience unemployment</b>	-0.1323	0.0015	-0.1354	0.0016	-0.0185	0.0034	-0.0134	0.0035	-0.0423	0.0039
<b>Education age interaction</b>	0.0011	0.0001	0.0010	0.0001	0.0009	0.0002	0.0010	0.0002	0.0012	0.0002
<b>Have child</b>	0.0189	0.0116	0.0281	0.0117	0.0169	0.0146	0.0100	0.0152	0.0196	0.0174
<b>Other household income</b>										
<b>Education level</b>	-0.1214	0.0056	-0.1479	0.0058	-0.0500	0.0174	-0.0596	0.0184	-0.1058	0.0189
<b>Education Squared</b>	0.0013	0.0002	0.0025	0.0002	0.0012	0.0006	0.0015	0.0006	0.0025	0.0007
<b>Inverse Mills Ratio</b>										
<b>Wage Expectation Correction</b>			-0.0002	0.0000			0.0000	0.0000		
<b>(Random Coeff sd)</b>							0.0005	0.0000		
<b>Intercept (Random sd)</b>					1.5028	0.0346	1.4760	0.0361		
<b>Control for Health</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Control for Sector</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Control for Year effect</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Control for Household Type</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Control for Cohort Effect</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

To conclude, the estimation results from the wage models are in line with expectation. Among all the specifications, the FEVD estimator appeared to fit the data the best.

The estimation results for the labour supply models are extensively examined in this study. Tables 5 and 6 present the coefficients estimated. Five labour supply models for both men and women are compared in the tables: the FEVD under certainty and uncertainty, the mixed fixed and random coefficients model under certainty and uncertainty, and the extended Heckman under certainty. The estimated coefficients are in general significant and stable across models.

The estimated wage elasticity is highly significant and stable for both men and women across models. Men record on average, in absolute value, a higher elasticity than women do. The highest wage elasticity is found under certainty when estimated using FEVD. Assuming that the effect of wage varies across individuals dampens the magnitude of the effect. Assuming uncertainty, the effect increases by 2 percentage points. The Heckman model provides higher wage elasticity estimates than the random coefficient model.



As expected, the wage expectation correction has a small but significant negative effect on labour supply in both the FEVD and the random coefficient models. Incorporating that people adapt their expectations in a heterogeneous manner dampens the average impact of the wage expectation correction.

With regard to age, the labour supply profile of men illustrates a higher growth rate of hours worked at younger ages than the profile for women. The rate of growth in hours worked, however, reduces more rapidly toward later ages for men than for women. When the age effect is allowed to vary across individuals, its effect reduces in magnitude compared to the FEVD variant. The Heckman model provides the highest estimates for the age effect.

The estimates for the return to experience (full time and part-time) are similar across models: all key variables are significant, have the same shape and similar magnitudes. For men, the cumulated work experience has a negative effect on the lifetime labour supply response, which can be explained by the age-labour supply profile. For women, the cumulated work experience has a small positive effect on the labour supply response. Unemployment experience has a negative effect for the labour supply of both men and women, and the effect is higher in absolute value for women.

The presence of children in the family has a negative effect on the labour supply of women, whereas for men the effect is not significant. The effects are stable across models. For women, the estimates of the inverse Mills ratios differ greatly across models. The estimate of the Heckman model is similar to the random coefficient model under uncertainty. To conclude, the estimates are stable across the different model specifications and show on average a high level of significance. The FEVD variant of the model fits the best.

### 1.V.2 Simulation Results

The simulation exercise evaluates the model's out-of-sample prediction performance. Each model is estimated on the same dataset and is used to predict the labour supply in year 2004 to 2005. The simulation uses a simple dynamic microsimulation framework, where only crucial demographic variables are updated. The performance of the simulation is judged according to the distance between the actual value and the predicted value. Table 7 gives a general overview of each model's performance using this indicator. Both simulated and actual values are coded in logarithm scale.

**Table 7 Simulation Models Comparison**

<i>Simulation Residual (logarithm)</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Heckman Extension Model	-0.354	0.432	-3.728	0.732
FEVD Fix coefficient with certainty	-0.027	0.444	-3.673	2.020
FEVD Fix coefficient with uncertainty	-0.031	0.441	-3.626	1.988
Mixed coefficients with certainty	-0.039	0.888	-8.364	8.936
Mixed coefficients with uncertainty	0.465	5.744	-50.933	55.689

The errors in the simulation come from three sources. The first source is the selection model, common to all models. The second source is the continuous labour supply models, which cover five different model setups. The last source of errors is the wage estimation, as in the models under uncertainty, the residual of the wage equation is crucial. In general, the FEVD fixed coefficients under certainty and uncertainty perform the best in terms of mean value and standard deviation of the simulated residuals. In terms of unbiasedness, the two

FEVD estimated models and the mixed coefficients model under certainty perform better than the Heckman extension model. The mixed coefficients estimation, however, can be significantly affected by outliers and has a large standard error for the random intercepts, which may lead to errors in the simulation.

Figure 2 sorts the observations in terms of the absolute difference between simulated and actual values. Since the wage expectation adjustment (uncertainty) term does not play a major role numerically as suggested in the estimation result, models with uncertainty correction have a very similar curve compared with the ones without, provided same estimation method is used. As a result, only 3 models are selected present in Figure 2. The figure shows that the models estimated with FEVD method performs the best in terms of the percentage of observations with the prediction error of less than 20 hours, while the Heckman extended model performs the worst judged by this criteria. This suggests that in a simulation study where the absolute error is a key factor, the proposed models estimated with FEVD.

**Figure 2 Simulation Residual for all three models 2004-2005**

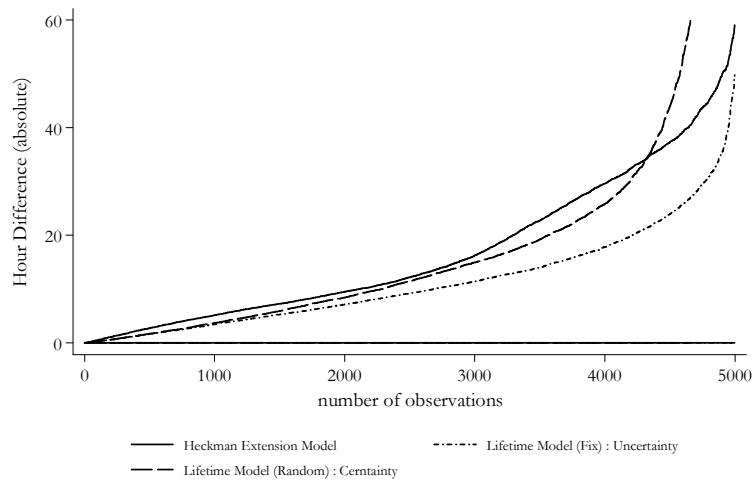
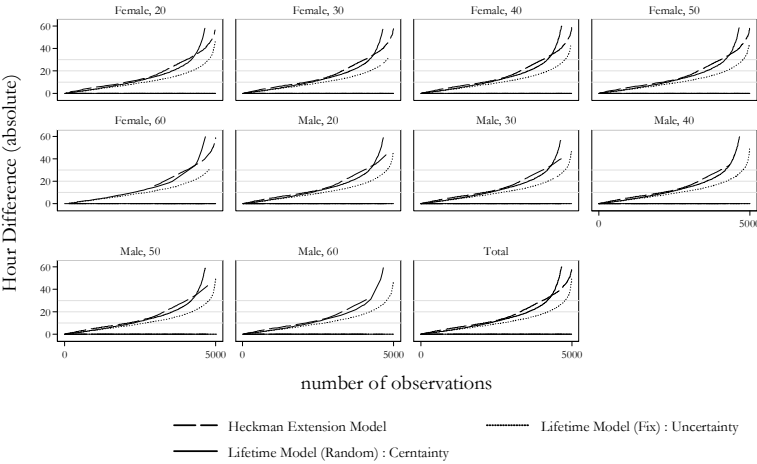


Figure 3 extends the previous graph by showing the results for different groups, whereas Figures 5.4 and 5.5 show the distribution of the simulation for different groups. The simulation shows that Heckman model is biased for 2004 and it continues to get worse in 2005. This bias may come from the biased estimation of time-invariant characteristics. The model estimated with mixed coefficients, although the mean predicted value is closer to the actual ones, it has a much larger standard error in the prediction. This finding is consistent with Figure 5, which shows that the mixed coefficients method has one of the most unbiased performances.

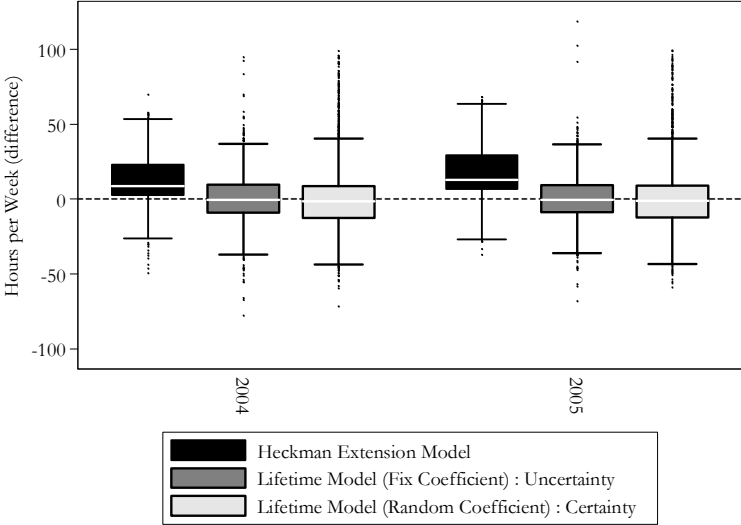
Figure 3 Simulation Residual for all three models 2004-2005



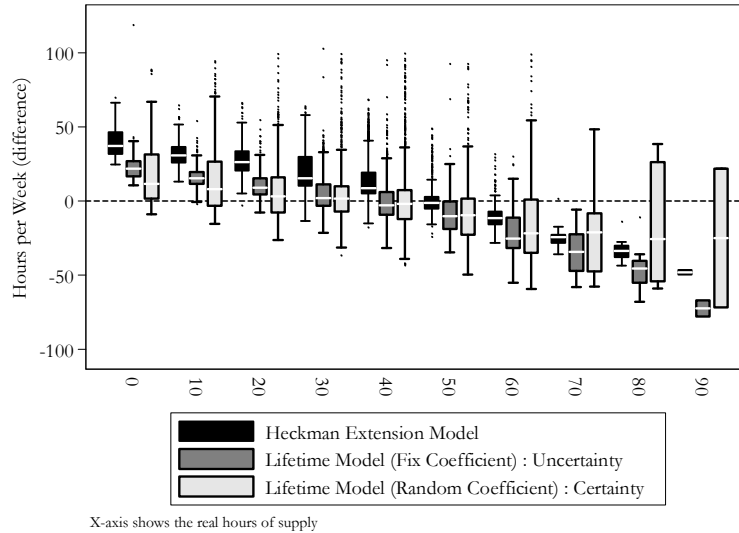
Graphs by Sex and Age Group (10 Years Interval)

Overall, the FEVD variant of the model has the best simulation result according to the validation tests. The uncertainty extension, however, does not seem to have a large impact, this may partially be due to the error from the wage estimation. The mixed coefficient estimation seems to handle the heterogeneity better than the other models, but is less than ideal for simulations.

Figure 4 Simulated Value Distributions by year



**Figure 5 Simulated Value Distributions by real value of working hours**



## VI. Conclusion

This chapter develops a structural lifetime model for estimating and simulating continuous labour supply. The model is consistent with the lifetime economic theory and is able to capture the individual heterogeneity to a larger extent than the existing labour supply models by using more refined estimation techniques, including fixed effect vector decomposition (FEVD) and the mixed coefficients estimation method. In addition, one variant of the model loosens the certainty assumption in the life cycle modelling. Instead, individuals are assumed to adjust their labour supply behaviours based on the differences between expected and actual earnings.

The chapter compares different combinations of the model specifications and estimation methods, as well as the standard random effects model (Heckman) for their simulation performances. In a simple simulation presented in this paper, models were estimated with different combinations of estimation techniques and uncertainty correction term. The results suggest that the models estimated using the FEVD method has the highest prediction accuracy judged by the mean error of simulation.

While the expectation correction introduced is also found to be significant in the estimation, it is found to be less important in the simulation exercise due to its relatively small coefficient and the potential wage estimation errors. When estimated using the FEVD method, the lifetime labour supply model developed in this study outperforms the Heckman panel extension model in the simulation by all indicators calculated. In practice, the models presented in this study could potentially benefit the microsimulation models where continuous labour supply models are used.

## References

- Altonji, J. G., 1986. Inter-temporal Substitution in Labor Supply: Evidence from Micro Data. *The Journal of Political Economy*, 94(3), S176-S215.
- Blundell, R., & MaCurdy, T. E., 1998. *Labor Supply: A Review of Alternative Approaches*: Institute for Fiscal Studies.
- Bover, O., Muellbauer, J. and Murphy, A., 1989. Housing, Wages and UK Labour Markets. *Oxford Bulletin of Economics and Statistics*, 51: 97–136.
- Card, D., 1991. Inter-temporal labour supply: an assessment. *Advances in Econometrics: Sixth World Congress*.
- Galler, H., 1996. *Microsimulation of pension reform proposal: Modelling the earning of couples in Harding A. Microsimulation and Public Policy*, New York, US: North-Holland.
- Haan, P., 2004. *Discrete Choice Labor Supply: Conditional Logit Vs. Random Coefficient Models*: DIW.
- Hsiao, 2003. *Analysis of Panel Data*, 2nd edition, Cambridge University Press
- MaCurdy, T. E., 1980. *An Empirical Model of Labor Supply in a Life Cycle Setting*: NBER.
- MaCurdy, T. E., 1983. A Simple Scheme for Estimating an Inter-temporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty. *International Economic Review*, 24(2), 265-289.
- MaCurdy, T. E., 1985. Interpreting Empirical Models of Labor Supply in an Inter-temporal Framework with Uncertainty. *Longitudinal Analysis of Labor Market Data*, 111-155.
- Medoff, J. L., & Abraham, K. G., 1981. Are Those Paid More Really More Productive? The Case of Experience. *The Journal of Human Resources*, 16(2), 186-216.
- Plumper T., Troeger V. E., 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15, 124–139
- Pudney S., 1992. "Dynamic simulation of pensioners' incomes: methodological issues and a design for a computer model for Great Britain", Cambridge.
- Rabe-Hesketh, S., & Skrondal, A., 2005. *Multilevel and Longitudinal Modelling using Stata*: Stata Press.
- Rendel, U., & Buechel, F., 1994. *Tests for non-ignorable panel attrition and their application on wage estimation from the German Socio Economic Panel*. Berlin: Deutsches Institut fuer Wirtschaftsforschung.
- Van Soest, A., 1995. Structural Models of Family Labor Supply: A Discrete Choice Approach. *The Journal of Human Resources*, 30(1), 63-88.
- Vella, F., 1998. Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources*, 33(1), 127-169.