

Löffler, Max; Peichl, Andreas; Siegloch, Sebastian

Working Paper

Structural labor supply models and wage exogeneity

ZEW Discussion Papers, No. 14-040

Provided in Cooperation with:

ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Löffler, Max; Peichl, Andreas; Siegloch, Sebastian (2014) : Structural labor supply models and wage exogeneity, ZEW Discussion Papers, No. 14-040, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, <https://nbn-resolving.de/urn:nbn:de:bsz:180-madoc-369019>

This Version is available at:

<https://hdl.handle.net/10419/98179>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Discussion Paper No. 14-040

Structural Labor Supply Models and Wage Exogeneity

Max Löffler, Andreas Peichl,
and Sebastian Siegloch

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 14-040

Structural Labor Supply Models and Wage Exogeneity

Max Löffler, Andreas Peichl,
and Sebastian Siegloch

Download this ZEW Discussion Paper from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/dp14040.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other
economists in order to encourage discussion and suggestions for revisions. The authors are solely
responsible for the contents which do not necessarily represent the opinion of the ZEW.

Structural Labor Supply Models and Wage Exogeneity*

Max Löffler Andreas Peichl Sebastian Siegloch

This version: June 24, 2014

Abstract. There is still considerable dispute about the magnitude of labor supply elasticities. While differences in micro and macro estimates are recently attributed to frictions and adjustment costs, we show that relatively low labor supply elasticities derived from microeconomic models can also be explained by modeling assumptions with respect to wages. Specifically, we estimate 3,456 structural labor supply models each representing a plausible combination of frequently made choices. While most model assumptions do not systematically affect labor supply elasticities, our analysis shows that the results are very sensitive to the treatment of wages. In particular, the often-made but highly restrictive independence assumption between preferences and wages is key. To overcome this restriction, we propose a flexible estimation strategy that nests commonly used models. We show that loosening the exogeneity assumption leads to labor supply elasticities that are much higher.

JEL Classification: C25, C52, H31, J22

Keywords: labor supply, elasticity, random utility models, wages

*Löffler is affiliated to the Centre for European Economic Research (ZEW), Mannheim and the University of Cologne (loeffler@zew.de); Peichl to ZEW, University of Mannheim, CESifo, ISEI and IZA (peichl@zew.de); Siegloch to IZA, ZEW and University of Mannheim (siegloch@iza.org). The authors thank Rolf Aaberge, Felix Bierbrauer, Richard Blundell, John Dagsvik, Alan Duncan, Peter Haan, Jan Kabátek, Nico Pestel and Eric Sommer, as well as seminar participants at CPB, IIPF, IZA, Young Scholar GSOEP Symposium, Statistics Norway and Verein für Socialpolitik for helpful comments and suggestions on earlier versions.

1 Introduction

Knowing the true size of labor supply responses has important implications for welfare analysis (Eissa et al., 2008) and optimal taxation (Diamond and Saez, 2011, Immervoll et al., 2011). One of the most topical questions in the long and comprehensive literature on labor supply behavior is why macro elasticities are (substantially) larger than micro ones (Chetty et al., 2011, Keane and Rogerson, 2012). While older explanations focus on the use of representative agents and aggregation difficulties (Blanchard, 2007) or social multipliers (Alesina et al., 2006), more recent studies attribute differences to costs and frictions in labor supply adjustments, which standard microeconomic research designs can only imperfectly account for (Chetty, 2012). In principle, structural micro-level labor supply models should be able to capture these adjustment frictions (Aaberge et al., 1995, van Soest, 1995, Blundell et al., 2000).¹ However, even these widely used models typically produce labor supply elasticities well below those found in macro studies, which immediately raises the question of why this is the case.

In this paper, we aim to answer this question by thoroughly scrutinizing state-of-the-art micro labor supply models and their functioning. More specifically, in the first part of the paper, we test whether the numerous modeling choices and assumptions to be made when setting up structural labor supply models affect estimated elasticities.² We check the internal validity of these models by running controlled experiments: we set up and estimate 3,456 models each representing a different (plausible) combination of commonly made modeling assumptions using two different micro datasets—one for Germany and one for the US. Based upon the estimation results, we gather insights into how robust the statistical fit of the models and the estimated labor supply elasticities are with respect to the underlying assumptions.

Our results show that the models' predictions do not depend on the specification of the functional form or the inclusion of observed and unobserved preference heterogeneity, hours restrictions or stigma costs of welfare participation. However, we find that the treatment of wages in the estimation procedure is crucial. For instance, the choice between predicting wage rates for the full sample or only for non-workers with missing wage information—both procedures are often used in the literature—may more than double the estimated labor supply elasticities. We conclude that the attention of previous sensitivity analyses has been mainly concentrated on less important factors while the main driving forces have been neglected, i.e. the interactions

¹While recent research in labor economics typically relies on quasi-experimental methods to identify causal effects of reforms, structural models are still necessary for policy analysis and especially out of sample predictions (e.g., the ex-ante evaluation of a tax reform).

²Note that structural models are sometimes criticized for the large number of assumptions to be made and the even larger number of possible combinations of these assumptions (Keane, 2010).

between wages, working hours and preferences. This finding is even more relevant given that most existing models (implicitly) assume exogeneity between the wage equation and the labor supply decision, which is naturally quite a restrictive assumption.³

Against this backdrop, in the second part of the paper, we propose a novel, flexible estimation strategy to relax the strict exogeneity assumptions regarding wage rates and work preferences. Estimation results show that there is indeed strong correlation between both preferences and wages, as well as wages and hours of work. For instance, wages are lower for part-time compared to full-time jobs and individuals with higher wages are found to have higher preferences for leisure. The usual procedure to estimate wages in the first step and assume a fixed wage rate (independent of working hours) for every individual in the labor supply estimation ignores these correlation patterns and drives the estimated elasticities towards zero. In our preferred model, estimated labor supply elasticities are more than twice as high compared to conventional models assuming zero correlation between work preferences and wages (0.6 instead of 0.25). We conclude that the standard approach neglects important factors that determine a household's response to wage changes, which yields too low micro elasticities of labor supply. Consequently, part of the unexplained difference between macro and structural micro elasticities may be due to model specification errors regarding the wage treatment.

In this paper, we make three important contributions to the literature on labor supply estimation. First, there is little evidence on the functioning of structural labor supply models in general. Moreover, if such studies exist, different models are not estimated on the same dataset.⁴ In that respect, we run a controlled meta-analysis, isolating the impact of the model assumptions on estimation outcomes. Second, our analysis points to a hitherto neglected factor that strongly influences the estimated labor elasticities: we show that the treatment of wages in labor supply estimations, which is rarely theoretically motivated nor subject to robustness checks, crucially affects the estimation results. In particular, we demonstrate that the commonly assumed exogeneity between wage rates and labor supply decisions plays an important role. To tackle this issue, we propose a novel estimation strategy that relaxes this assumption and additionally allows for correlation between work preferences and wages. There-

³Only little effort has been made thus far in the context of discrete choice labor supply models to overcome this assumption. Aaberge et al. (1995), Breunig et al. (2008) and Blundell and Shephard (2012) estimate preferences and wages simultaneously, in part also allowing for some correlation. We discuss the differences to our approach, which is more general, in detail in Section 5.

⁴Existing surveys and meta-analyses focus on either the principles of alternative estimation strategies (Blundell and MaCurdy, 1999, Evers et al., 2008) or cross-country comparisons of empirical findings (Bargain et al., 2014). Robustness checks in previous studies usually limit themselves to small deviations in one or only few of the numerous modeling assumptions.

fore, our approach is more flexible than previous models, while at the same time it nests previous models and allows testing the assumption of zero correlation. Third, our findings of significant correlations and resulting higher labor supply elasticities have important policy implications as labor supply elasticities are key parameters when evaluating or designing optimal tax benefit policies. For instance, Diamond and Saez (2011) use an elasticity of 0.25, which is close to our estimate for the restrictive model, to derive an optimal top marginal tax rate of 72.7%. However, an elasticity of 0.6, as found in our most flexible model, reduces the optimal tax rate to 52.6%, bringing it closer to actually observed values.

The remainder of this paper is organized as follows. Section 2 presents the general modeling framework and a short overview of the existing literature. Section 3 provides information on the used data and the modeling of the tax and benefit system. In Section 4 we conduct our analysis of modeling assumptions and present first results. The new flexible estimation approach is introduced in Section 5, before Section 6 concludes.

2 Model and existing literature

The use of structural discrete choice labor supply estimations has become a standard procedure in the empirical analysis of labor supply for both econometricians and policy makers (Bargain and Peichl, 2013). The first generation of labor supply models relied on the assumption that the household's utility is maximized over a continuous set of working hours—known as *Hausman approach* (see Hausman, 1981). This procedure has proven somewhat cumbersome when the budget set is non-convex, which will often be the case in presence of the complicated tax and benefits systems in most countries. Moreover, it has been shown that the estimated models are very sensitive to the underlying wages (Ericson and Flood, 1997, Eklöf and Sacklén, 2000).

As the consistent estimation of this type of model relies on rather restrictive a priori assumptions (see, e.g. MaCurdy et al., 1990, or Bloemen and Kapteyn, 2008, for details), it has become increasingly popular to model the labor supply decision as choice between a (finite) set of utility levels instead of deriving the marginal utility. Starting with the works by Aaberge, Dagsvik and Strøm (1995), van Soest (1995) and Hoynes (1996), a wide range of different empirical specifications of these *discrete choice models* has been applied. Comparing different levels of utility avoids the cumbersome maximization process of Hausman-type models. We focus our analysis on the discrete choice approach, given that it has become the standard procedure across the literature.

2.1 General model

Structural labor supply estimations build on the assumption of the well-known neo-classical labor supply model that decision makers maximize their utility by choosing the optimal amount of hours of work (or the optimal job, more generally). As higher working hours increase consumption but reduce leisure, households face a trade-off between these two goods. Stated mathematically for individual n :

$$\max_j U(C_{nj}, L_j, \epsilon_{nj}) = \max_{j \in J_n} U(f\{w_{nj}h_j | x_{nj}, I_n\}, T - h_j, \epsilon_{nj}) \quad (1)$$

where leisure L_j is denoted as difference between total time endowment T and working hours h_j . Consumption C_{nj} depends on working hours, hourly wage rate w_{nj} , non-labor income I_n , household and job characteristics x_{nj} and the tax benefit function f . Individual n faces the decision between a set of jobs J_n with working hours h_j and wages w_{nj} , including non-participation denoted as $j = 0$ (with $h_0 = 0$ and $w_{n0} = 0$). Most models in the literature make the quite restrictive assumption that wages are individual-specific and do not vary across alternatives.⁵ We assume a static context, which implies that consumption equals disposable income as there is no future utility from saving.

The true utility is only partly observable to the researcher while other components ϵ_{nj} are latent. Thus, we can write the utility of individual n choosing job type j as:

$$U(C_{nj}, L_j | x_{nj}, \beta_n, \gamma_n) = \varphi(C_{nj}, L_j | x_{nj}, \beta_n, \gamma_n) + \epsilon_{nj} \quad (2)$$

The observed part $\varphi(C_{nj}, L_j | x_{nj}, \beta_n, \gamma_n)$ is determined by consumption and leisure, characteristics x_{nj} , individual preferences β_n and labor market conditions γ_n that capture the availability of job type j . One could think of these labor market characteristics γ_n as measuring individual-specific fixed costs, search costs for part-time jobs or rigidities regarding working hours, for example. The unobserved taste variation ϵ_{nj} is assumed to be i.i.d. and follow the extreme value type I distribution with cumulative distribution function $F(\epsilon) = \exp(-\exp(-\epsilon))$. McFadden (1974) has shown that the probability of individual n choosing a job of type i is subsequently given by:

$$P(U_{ni} > U_{nj}, \forall j \neq i | x_n, \beta_n, \gamma_n) = \frac{\exp(\varphi\{C_{ni}, L_i | x_{ni}, \beta_n, \gamma_n\})}{\sum_{s \in J_n} \exp(\varphi\{C_{ns}, L_s | x_{ns}, \beta_n, \gamma_n\})} \quad (3)$$

⁵An exception are the models by Aaberge, Dagsvik and Strøm (1995) and follow-ups. We further loosen this restriction in Section 5 when estimating wages and preferences jointly.

Assuming that individuals take labor market conditions as given, we can rewrite:

$$P(U_{ni} > U_{nj}, \forall j \neq i | \mathbf{x}_n, \boldsymbol{\beta}_n, \gamma_n) = \frac{\exp(v\{C_{ni}, L_i | \mathbf{x}_{ni}, \boldsymbol{\beta}_n\}) g(i | \mathbf{x}_{ni}, \gamma_n)}{\sum_{s \in J_n} \exp(v\{C_{ns}, L_s | \mathbf{x}_{ns}, \boldsymbol{\beta}_n\}) g(s | \mathbf{x}_{ns}, \gamma_n)} \quad (4)$$

with $v(C_{nj}, L_{nj})$ as systematic utility function and $g(j)$ as frequency of feasible jobs with type j . In words, the individual choice probability is given as the systematic utility part weighted by the availability of jobs with type j . In the following, we discuss the specification of $v(\cdot)$ and $g(\cdot)$ and the estimation procedure.

2.2 Estimation

Econometrically, the discrete choice approach boils down to the representation of the labor supply decision in a random utility model. In the very basic model, the theoretical setup implies that the household's decision satisfies the Independence of Irrelevant Alternatives (IIA) property (Luce, 1959). In other words, the preference between two alternatives does not depend on the presence of a third one. While this assumption may seem rather restrictive at first glance, Dagsvik and Strøm (2004) and Train (2009) show that it is well in line with economic intuition and even less restrictive than the necessary assumptions to estimate continuous hours models. However, the IIA assumption is no longer needed as soon as additional random effects are incorporated in the model (see Section 2.3).

Identification. As in every structural estimation problem, it is crucial to impose a specific functional form for both $v(C_{nj}, L_{nj})$ and $g(j)$ to obtain consistent estimates of $\boldsymbol{\beta}_n$ and γ_n . Van Soest et al. (2002) show that semi-parametric specifications also yield consistent results. We further discuss different assumptions that are frequently used below. As consumption is a function of working hours and thus leisure, identification of preference parameters relies on (a) the variation in working hours h_j , hourly wages w_{nj} , non-labor income I_n and other characteristics \mathbf{x}_{nj} and (b) the fact that the tax function $f(w_{nj}h_j, I_n)$ is highly non-linear in h_j and w_{nj} . This also implies that labor market conditions γ_n can only be separated and identified on the assumption of a specific functional form (Dagsvik and Strøm, 2006).

In addition to this, the vast majority of the literature also assumes that preferences $\boldsymbol{\beta}_n$ and labor market conditions γ_n may depend on individual characteristics, but are independent of the wage rate w_{nj} . Thus, it is assumed that:

$$\text{Corr}(\boldsymbol{\beta}_n, w_{nj} | \mathbf{x}_{nj}) = 0 \quad \text{Corr}(\gamma_n, w_{nj} | \mathbf{x}_{nj}) = 0 \quad (5)$$

The main reason for this assumption is that it reduces the computational burden substantially and makes the estimation more convenient. However, assuming exogeneity seems quite restrictive, as unobserved ability, for instance, is most likely positively correlated with both wage rates and work preferences. Therefore, we test and relax this assumption in Section 5.

In order to estimate the preference coefficients, one has to evaluate both functions v and g for every household $n = 1, \dots, N$ and every choice category within the choice set J_n . Given the different income levels, the model can be estimated via maximum likelihood. The derivation of the (log)-likelihood function is straightforward (McFadden, 1974). However, some modeling assumptions have to be made, as well as several possible extensions to this simple setup.

2.3 Model setup

Choice set. The first decision in the estimation relates to the construction of the choice set (see Aaberge et al., 2009, for a detailed discussion of this issue). Most authors simply pick a set of representative levels of hours of work and assume (small) identical choice sets for the whole population. In our analysis, we follow the literature and assume that households with a single decision maker face seven possible labor supply states, i.e. either non-participation or working 10, 20, 30, 40, 50 or 60 hours per week. Couple households are assumed to face 7^2 alternatives.⁶

Functional form of the systematic utility. As the discrete choice approach relies on the comparison of different utility levels, it is crucial to determine the form of the systematic utility function. In theoretical terms, the function v represents the direct utility function of the household. Most applications rely on either a translog, a quadratic or a Box-Cox transformed utility specification. However, several other choices are possible.

Heterogeneity in preferences. Observed heterogeneity in the labor supply behavior can be rather easily introduced in the context of structural labor supply models. The preference coefficients of the direct utility function are usually interacted with some observed household characteristics, such as age or the presence of children, as taste shifters.

⁶Note that the results are generally not sensitive to the number of choices (e.g., 4 vs. 7 vs. 13) or the exact value assigned to each category – see, e.g., Bargain et al. (2014).

Moreover, including also unobserved heterogeneity overcomes the IIA assumption as it allows for unobservable variation in preferences between choice alternatives. There are two main ways to do so: in most applied works, either a *random coefficient model* (van Soest, 1995) or a *latent class model* (Hoynes, 1996) is assumed. While the former assumes a set of coefficients to be (multivariate) normally distributed, the latter assumes a set of discrete mass points for the estimated coefficients. Keane and Wasi (2012) discuss the performance of both approaches. We focus on the random coefficient approach as it has become standard across the literature.

Welfare stigma and benefit take-up. While the model as described thus far assumes that households only build their preferences with respect to the levels of consumption and leisure, their utility may also depend on the *source* of income. For example, the participation in welfare programs may be connected to an unobservable stigma that affects the household's utility and prevents some households from taking up benefits (Moffitt, 1983). In the discrete choice context, this can be incorporated by accounting for the potential disutility from welfare participation and expanding the choice set such that the household explicitly chooses between benefit take-up and non-participation (Hoynes, 1996, Keane and Moffitt, 1998).

Fixed costs and hours restrictions. Moreover, van Soest (1995) argues that working part-time could also be connected with an unobservable disutility, because part-time jobs may exhibit higher search costs. Euwals and van Soest (1999) extend this idea by introducing fixed costs of work, which have since been used in several applications. While both approaches help to explain the observed labor market outcomes, their rational remains rather ad hoc. Aaberge et al. (1995) provide a more convenient theoretical framework that delivers a structural interpretation of fixed costs and the utility connected to certain hours alternatives. In their model, households choose between (latent) job offers that not only differ regarding the working hours, but also in terms of availability, wages and non-monetary attributes.

2.4 Wage imputation procedure

In addition to the specification of the utility function, there are important modeling assumptions regarding the wage imputation. In order to calculate the disposable income for the different choice alternatives, one needs information on the hourly wage rates. While for actual workers the wage rate can be calculated by gross earnings and hours of work (we use standardized working hours to reduce the potential division bias, see Borjas, 1980, and Ziliak and Kniesner, 1999, for a discussion), the wage

information is typically missing for non-workers. The first decision is how to deal with missing wages in the estimation process. In practice, wages are either estimated beforehand and treated as given within the estimation of the labor supply model or wages and preferences are estimated jointly. In addition, one has to decide whether the estimated wage rates are used only if wages are not observed or for the full sample (see MaCurdy et al., 1990, for a discussion of the advantages and disadvantages of both approaches). In either case, one can ignore or explicitly include potential sample selection issues in the observed wages.

After estimating the wage equation, another important question is whether the potential errors in the wage rate prediction are incorporated in the labor supply estimation. Especially when using predicted wages for the full sample, the “new” distribution of wages will typically have a significantly lower variance and the predicted wage will differ considerably from the observed one, at least for some workers. Thus, ignoring the error when predicting wage rates, which is still done in practice, leads to inconsistent estimates. The standard procedure to incorporate wage prediction errors is to integrate over the whole estimated wage distribution and thus integrating out the wage prediction error during the estimation process (van Soest, 1995). One approximation used in some applications is to simply add a single random draw to the predicted wage rates (Bargain et al., 2014). While this procedure lacks a theoretical rationale, it substantially reduces the computational burden of the estimation.

2.5 Estimation approach

The named extensions especially regarding the inclusion of unobserved heterogeneity and the incorporation of wage prediction errors complicate the estimation procedure and lead to the more general representation as *mixed logit model* (Train, 2009). Taking the most general specification as reference, the likelihood function can be written as:

$$L = \prod_{n=1}^N \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\exp(v_{ni} \{ \cdot | \hat{w}_{ni}, \beta_n \}) g(i | \gamma_n)}{\sum_{j \in J_n} \exp(v_{nj} \{ \cdot | \hat{w}_{nj}, \beta_n \}) g(j | \gamma_n)} f(\beta_n, \gamma_n) f(\hat{w}_n) d\beta_n d\gamma_n d\hat{w}_n \quad (6)$$

where $i \in J_n$ denotes the alternative chosen by individual n . The likelihood contributions not only depend on the systematic utility function, but also on the availability of the choice alternatives, denoted by $g(i)$. This setup implies that the availability of choice alternatives can be separated from the systematic utility, which is a reasonable assumption at least for labor markets in industrialized countries. As the preferences may also include unobserved heterogeneity, the probability that household n maximizes its utility at choice alternative i has to be integrated over the possible set

of coefficients (β_n, γ_n) . Similarly, the individual likelihood contributions have to be integrated over the range of possible wage predictions \hat{w}_{nj} . As both variables will typically not be uniformly distributed, the choice probability has to be weighted by the probability density of the random components.

The model as written down in equation (6) is very general and less restrictive than the conditional logit setup. In turn, it is no longer possible to find an analytical solution. Train (2009) proposes the use of maximum simulated likelihood methods instead. In order to retrieve the simulated likelihood, the double integral has to be approximated and averaged over $r = 1, \dots, R$ random draws from the distributions of (β_n, γ_n) and \hat{w}_{nj} . The simulated log-likelihood is subsequently given by:

$$\ln(SL) = \sum_{n=1}^N \ln \left(\frac{1}{R} \sum_{r=1}^R \frac{\exp \left(v_{ni} \left\{ \cdot \mid \hat{w}_{ni}^{(r)}, \beta_n^{(r)} \right\} \right) g \left(i \mid \gamma_n^{(r)} \right)}{\sum_{j \in J_n} \exp \left(v_{nj} \left\{ \cdot \mid \hat{w}_{nj}^{(r)}, \beta_n^{(r)} \right\} \right) g \left(j \mid \gamma_n^{(r)} \right)} \right) \quad (7)$$

When the number of draws goes to infinity, the simulated log-likelihood in (7) converges to the log-likelihood of the model denoted in (6). Instead of relying on conventional random draws, we approximate the likelihood function using pseudo-random Halton sequences. This reduces the number of draws needed to ensure stable results as Halton sequences cover the desired distribution more evenly (Train, 2009).⁷

2.6 Existing literature

Tables 1 and 2 provide an overview of the empirical specification of several popular models that have been applied in recent years. As one can see, mainly three utility functions have been used, i.e. either a translog, a quadratic or a Box-Cox transformed specification. As the Stone-Geary function can be interpreted as a simplification of the translog or the Box-Cox utility function, only the higher-degree polynomials used in van Soest et al. (2002) stand out from the list. Their approach can be seen as approximation to a non-parametric specification of the utility function. The inclusion of observed heterogeneity shows a similar picture. All studies allow for observed heterogeneity in the preferences for leisure, whereas fewer studies allow for preference heterogeneity regarding consumption. The evidence on unobserved heterogeneity is somewhat more mixed, just like the inclusion of heterogeneity in fixed costs and the potential stigma from welfare participation.

As working hours are typically concentrated in only few hours categories, most authors include in their models fixed costs of working, hours restrictions, or both.

⁷Details on the estimation procedure can be found in Löffler (2013).

Table 1: Different model specifications

Paper	Utility	Heterogeneity [*]		Welfare	
	Function	Observed	Unobs.	Stigma	Constraints
Aaberge et al. (1995, 2009)	Box-Cox	L	—	—	FC, HR
Aaberge et al. (1999)	Box-Cox	L, FC	—	—	FC, HR
Dagsvik and Strøm (2006)	Box-Cox	L, FC	—	—	FC, HR
Dagsvik et al. (2011)	Box-Cox	L, FC	—	—	FC, HR
Blundell and Shephard (2012)	Box-Cox	L, C, S, FC	C, S	Yes	FC
van Soest (1995)	Translog	L	$—/L^{\dagger}$	—	$—/HR$
Euwals and van Soest (1999)	Translog	L, FC	L	—	FC
van Soest and Das (2001)	Translog	L, FC	L	—	FC
Flood et al. (2004)	Translog	L, L^2, S	L, L^2, S	Yes	—
Haan (2006)	Translog	L, C	$—/C$	—	HR
Flood et al. (2007)	Translog	L, C, FC, S	L, C, FC, S	Yes	FC
Hoynes (1996)	Stone-Geary	L, S	L, S	Yes	$—/FC$
van Soest et al. (2002)	Polynomial	L	L	—	FC
Keane and Moffitt (1998)	Quadratic	L, S	L, S	Yes	—
Blundell et al. (1999, 2000)	Quadratic	L, C, FC	C, S	Yes	FC
Bargain et al. (2014)	Quadratic	L, C, FC	C	—	FC

^{*} L and C denote heterogeneity in preferences for leisure and consumption, respectively. S denotes the disutility from welfare participation. FC refers to fixed costs of working and HR to hours restrictions.

[†] Robustness checks and alternative model specifications are separated by slashes.

Table 2: Wage imputation methods

Paper	Estimation Approach	Sample Selection	Imputation	Prediction Error
Aaberge et al. (1995, 2009)	Simultaneous	—	Full sample	—
Aaberge et al. (1999)	Simultaneous	—	Full sample	—
Keane and Moffitt (1998)	Simult./Two step [*]	—	Non-workers	—
van Soest et al. (2002)	Simultaneous	—	Non-workers	Integrated out
Blundell and Shephard (2012)	Simult./Two step	—	Non-workers	Integrated out
van Soest (1995)	Two step	Yes	Non-workers	$—/Integrated$ out
Euwals and van Soest (1999)	Two step	Yes	Non-workers	Integrated out
Blundell et al. (1999, 2000)	Two step	Yes	Non-workers	Integrated out
van Soest and Das (2001)	Two step	Yes	Non-workers	Integrated out
Haan (2006)	Two step	Yes	Non-workers	—
Flood et al. (2007)	Two step	Yes	Non-workers	$—/Integrated$ out
Dagsvik et al. (2011)	Two step	Yes	Non-workers	—
Hoynes (1996)	Two step	Yes	Full sample	—
Flood et al. (2004)	Two step	Yes	Full sample	—
Dagsvik and Strøm (2006)	Two step	Yes	Full sample	Integrated out
Bargain et al. (2014)	Two step	Yes	Full sample	Random draw

^{*} Robustness checks and alternative model specifications are separated by slashes.

Fixed costs and hours restrictions can also be interpreted as measures for the availability of the respective choice alternatives (Aaberge et al., 2009). Less than half of the models explicitly allowed for stigma effects and non-take-up of welfare benefits. This is interesting due to the common finding that the benefit participation rate deviates substantially from full take-up. Thus, models that do not explicitly account for the potential disutility are expected to over-predict the number of recipients.

Less variation can be found in terms of the model's treatment of wages. While most studies estimate wages and the labor supply decision in a two-step procedure, only the models of Aaberge et al. (1995, and follow-ups), Keane and Moffitt (1998), van Soest et al. (2002) and Blundell and Shephard (2012) apply a simultaneous maximum likelihood procedure. In turn, these models neglect potential sample selection issues when estimating wages. As can be seen, there is no consensus in the literature whether predicted wages should be used only for individuals with unobserved wages or for the full sample, the advantage of the latter choice is there is eventually only one wage distribution. Regarding the handling of the wage prediction errors, it becomes increasingly common practice to incorporate and integrate out the errors during the estimation.

3 Data

The baseline estimations in this paper are performed on the German Socio-Economic Panel (SOEP), a representative household panel survey for Germany (Wagner et al., 2007). SOEP includes now more than 24,000 individuals in around 11,000 households. We use the 2008 wave of SOEP, which provides household data from 2008, as well as data on the labor supply behavior and incomes from the preceding year (i.e. the year before the Great Recession). We rely on the tax and transfer system of 2007, focusing our analysis on the working age population and thus excluding individuals younger than 17 or above the retirement age of 65 from our estimations. Our sample is further restricted to those households where at least one decision maker can adjust her labor supply. Therefore, we exclude households where all decision makers are self-employed (since it is difficult to measure true hours and wages for those), civil servants⁸ or in the military service. Moreover, our subsample includes some households with more than two adults, which mainly includes adult children living with their parents. We exclude these young adults from the estimation as it is unclear how their consumption and utility are determined (Dagsvik et al., 2011).

⁸Tenured civil servants cannot freely adjust the weekly working hours. Note that we keep all other public sector employees.

As the labor supply decision is known to be rather heterogeneous across population subgroups, we separate the sample into five distinct demographical subpopulations. The first two groups are defined as single men and single women with or without dependent children. Our estimation subsample contains 779 households with single males and 1,065 households with single females. In addition, we specify three different kinds of couple households. First, we define 688 couple households in which the male partner has a flexible labor supply but the female partner is inflexible (e.g. due to self-employment or exclusion restrictions regarding the age). Second, we have 1,042 couple households in which the male partner has an inflexible labor supply but the female partner is flexible. In order to model the household labor supply decision of these “semi-flexible” couple households, we assume that the flexible partner faces his or her labor supply decision conditional on the labor supply behavior of the inflexible partner. Third, our sample includes 3,099 couple households in which both partners are flexible regarding their labor supply behavior.

For the computation of consumption levels for the different choice categories, we rely on IZA’s policy simulation model IZAΨMOD (v3.0.0), which incorporates a very detailed representation of the German tax and benefit system (see Peichl et al., 2010, for a comprehensive documentation). Some of the estimated models would require applying the tax and benefit system for every possible wage rate for every household in every step of the numerical likelihood maximization, although doing so would slow down the estimation process substantially. To avoid this cumbersome procedure, we approximate the tax and benefit system by using a highly flexible second-degree polynomial that transforms monthly gross earnings into disposable income while controlling for a rich set of household characteristics, as well as all available sources of non-labor incomes. The resulting R^2 shows a very good fit of more than 99 % for all population subgroups but single women (only 97 % for them), which confirms that our approximation performs rather well. In order to also allow for unobserved tax determinants, we balance the predicted amounts of consumption by a single random draw for each household; otherwise, we would mistakenly reduce the variance in the consumption variable. The results are very much in line with those taking advantage of the full representation of the tax and transfer system, we are thus confident that the approximation does not affect our findings.

As a robustness check, we compare our results obtained with German data to results for the USA. For this, we use data from IPUMS-CPS which is an integrated dataset of the March Current Population Survey (CPS) for 2007. In order to calculate income and payroll taxes, we use NBER’s simulation model TAXSIM.

4 Empirical results

Despite the presence of some robustness checks in the literature (see tables 1 and 2), these checks are usually narrowed down to a small deviation in just one of the modeling assumptions. By contrast, Evers et al. (2008) and Bargain and Peichl (2013) perform meta-analyses of labor supply models comparing estimated labor supply elasticities for different countries and explain them mainly by study characteristics. In either case, it is difficult to draw detailed and general conclusions on the specification of discrete choice models from the reported results. We overcome these difficulties by estimating a large variety of different modeling assumptions in a controlled environment using the same data. The estimation results allow us to determine how sensitive the estimated outcomes are with respect to the specification and the wage imputation procedure used in the model.

4.1 Analysis setup

For our analysis, we combine frequently used modeling assumptions and estimate all sensible combinations of these specifications. We estimate 3,456 different model specifications for the five distinct labor supply types, which leads us to 17,280 maximum likelihood estimations. However, the sample of estimation results is reduced because not all models did converge in a reasonable time span as we applied an automatic routine to find initial values and estimate this large number of models. We drop from our analysis those estimation results that did not converge. Depending on the labor supply group we lose up to 6 percent of our sample and end up with 16,730 different estimation results.⁹

Table 3 shows the different specifications and the number of converged estimation results and reads as follows. We estimate 1,152 distinct models with a Box-Cox transformed utility specification for each of the five labor supply groups, although only 1,022 estimation results for single males and 1,132 for single females are included in our sample. Regardless of the functional form of the utility function, 1,152 of the estimated models neglect any kind of hours restrictions or fixed costs, 1,152 models include part-time restrictions and 1,152 models account for fixed costs of work.

In order to make the estimation results comparable across the different labor supply groups, we standardize the statistical fit and the estimated elasticities within a labor supply group. We subsequently pool the data and regress the estimation results on the different modeling assumptions (mainly represented as dummy variables). We

⁹We drop models that did not converge after 100 iterations of Stata's maximum likelihood implementation. Of course, more complex models take longer to converge. Apart from that, we do not find systematic effects of different types of assumptions on the probability of not converging.

Table 3: Estimated model combinations

Model Parameter	Option	N	Number of Converged Models*				
			SgM	SgF	CoM	CoF	CoMF
Utility function	Box-Cox	1,152	1,022	1,132	951	1,148	1,029
	Quadratic	1,152	1,152	1,151	1,152	1,133	1,152
	Translog	1,152	1,125	1,144	1,148	1,148	1,143
Welfare stigma	No	1,728	1,642	1,701	1,607	1,713	1,664
	Yes	1,728	1,657	1,726	1,644	1,716	1,660
Hours restrictions	—	1,152	1,091	1,141	1,040	1,131	1,109
	Fixed costs	1,152	1,064	1,137	1,061	1,149	1,063
	Part-time	1,152	1,144	1,149	1,150	1,149	1,152
Number of Halton draws	—	288	288	288	283	288	286
	10	1,584	1,440	1,564	1,429	1,559	1,456
	5	1,584	1,571	1,575	1,539	1,582	1,582
Observed heterogeneity	—	864	835	864	822	860	834
	in C only	864	827	862	834	861	822
	in L only	864	827	858	798	859	836
	in L, C	864	810	843	797	849	832
Unobserved heterogeneity	—	576	574	571	566	570	574
	in C only	864	863	853	846	862	863
	in L only	576	520	574	523	569	541
	in L, C	864	804	856	795	854	791
	with correl.	576	538	573	521	574	555
Wage imputation	Full sample	1,728	1,652	1,708	1,635	1,710	1,655
	Non-workers	1,728	1,647	1,719	1,616	1,719	1,669
Wage prediction error	—	1,296	1,217	1,293	1,219	1,291	1,245
	1 random draw	1,296	1,236	1,291	1,203	1,284	1,239
	Integrated out	864	846	843	829	854	840
Total		3,456	3,299	3,427	3,251	3,429	3,324

* Single males (females) are denoted by SgM (SgF). Couples where only the male (female) partner has a flexible labor supply are denoted by CoM (CoF). CoMF denotes fully flexible couples. The column N refers to the number of possible model combinations for each choice whereas the subsequent columns report the number of converged models.

measure the statistical fit by the Akaike Information Criterion (AIC) of the models. To retrieve (uncompensated) labor supply elasticities, we increase the own-wage rates by ten percent and simulate the labor supply reaction to this wage change.

4.2 Estimation results

The results of these meta-regressions can be found in table 4. As the dependent variables have been standardized, the coefficients are difficult to interpret. Our results show, e.g. that using a quadratic utility function increases the AIC by 12 % of a standard deviation in the sample and thereby worsens the statistical fit. These results have to be compared to a rather simple reference model using a translog utility

Table 4: Marginal impact of modeling assumptions (SOEP)

	Fit	10 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Quadratic	0.119*** (0.023)	0.124*** (0.028)	-0.015 (0.062)	0.004 (0.053)
Box-Cox	-0.020 (0.026)	0.116*** (0.040)	0.080** (0.035)	0.085** (0.034)
Welfare stigma	0.968*** (0.076)	0.045 (0.062)	0.065 (0.047)	0.065 (0.042)
Number of Halton draws	-0.010*** (0.001)	0.005 (0.004)	-0.003 (0.004)	-0.002 (0.004)
Hours restrictions				
Part-time restrictions	-1.647*** (0.082)	0.384*** (0.070)	0.105** (0.039)	0.152*** (0.042)
Fixed costs	-1.093*** (0.070)	0.481*** (0.067)	0.187*** (0.040)	0.238*** (0.041)
Observed heterogeneity				
in C only	-0.335*** (0.057)	-0.049 (0.035)	0.060** (0.022)	0.043* (0.023)
in L only	-0.381*** (0.061)	0.048 (0.038)	0.045** (0.021)	0.046* (0.023)
in C and L	-0.475*** (0.070)	0.016 (0.044)	0.012 (0.019)	0.013 (0.022)
Unobserved heterogeneity				
in C only	0.005 (0.014)	-0.006 (0.023)	-0.059* (0.032)	-0.051 (0.030)
in L only	0.005 (0.013)	-0.081*** (0.023)	-0.029 (0.028)	-0.037 (0.027)
in C and L	-0.041*** (0.013)	-0.037 (0.024)	-0.069** (0.027)	-0.064** (0.026)
in C and L (with correl.)	-0.119*** (0.016)	-0.082** (0.034)	-0.102*** (0.034)	-0.101*** (0.033)
Wage imputation				
Full sample, no correction	-0.811*** (0.119)	2.121*** (0.094)	2.235*** (0.091)	2.240*** (0.086)
Full sample, error integrated out	-0.530*** (0.048)	1.399*** (0.119)	1.385*** (0.123)	1.406*** (0.124)
Full sample, 1 random draw	-0.104** (0.049)	0.071 (0.062)	0.131 (0.093)	0.121 (0.088)
Non-workers, error integrated out	0.000 (0.067)	0.048 (0.063)	0.040 (0.041)	0.041 (0.041)
Non-workers, 1 random draw	0.070 (0.056)	-0.230*** (0.038)	-0.232*** (0.035)	-0.235*** (0.037)
Constant	1.004*** (0.121)	-0.939*** (0.094)	-0.678*** (0.087)	-0.726*** (0.087)
Labor supply types	Yes	Yes	Yes	Yes
Observations	16730	13219	13219	13219
R^2	0.854	0.849	0.870	0.881

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model—the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

function, neglecting observed and unobserved heterogeneity in preferences as well as fixed costs of working, hours restrictions or any stigma from welfare participation. In this reference model, we use observed wage rates for actual workers and predict wages for non-workers without incorporating the wage prediction error in the labor supply estimation. We find many statistically significant relationships. However, the presented standard errors are not bootstrapped, given that this would render our sensitivity analysis computationally infeasible. Bootstrapped standard errors would be substantially larger than those presented. As the coefficients are measured in standard deviations, only those of at least one standard deviation (in absolute values) are also economically interesting. We summarize the key findings below.

Goodness of fit. Although the statistical fit is usually not the main outcome of interest, our results show several interesting patterns for future applications (see table 4). First, the choice of the utility function does not systematically improve or worsen the statistical fit. Our analysis confirms the usual finding that the implementation of hours restrictions, fixed costs and observed preference heterogeneity clearly help to explain the labor supply choices. Estimating random coefficients models that also allow for unobserved heterogeneity yields little value-added compared to the computational burden of the estimation. The results regarding the wage imputation show that these specification decisions also affect the statistical fit of the model substantially. Predicting wages not only for non-workers but also for the full sample improves the fit significantly. However, this is unsurprising as it demonstrates how much of the variation in the data is lost by using predicted instead of actual wages for the full sample when not accounting for errors in the wage rate prediction.

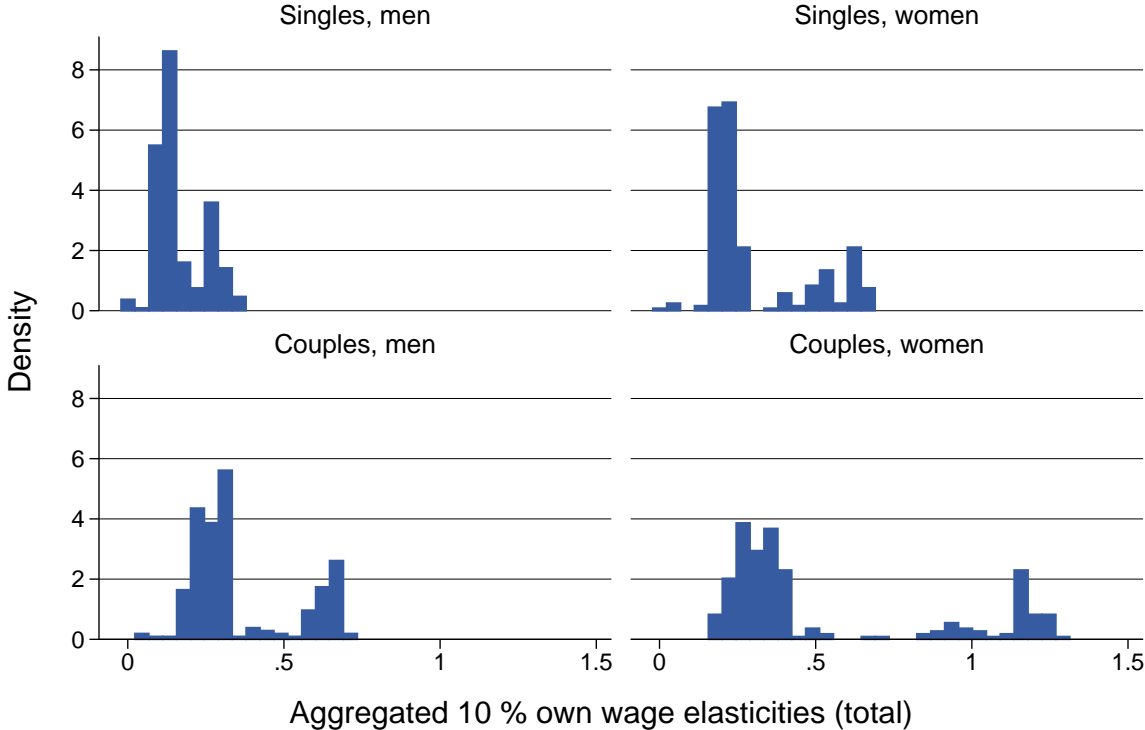
More generally, our results show that apart from the implementation of fixed costs or hours restrictions, there is hardly a single modeling assumption that guarantees a good fit. Instead, several small issues help to explain the observed labor market outcomes and add up to a good fit.

Labor supply elasticities. Even more important than the statistical fit is whether specific modeling assumptions systematically influence the out-of-sample predictions when simulating policy or wage changes. Figure 1 shows the distribution of labor supply elasticities across the converged models for the four labor supply types.¹⁰ The graph shows considerable variation across the different modeling setups (within a specific labor supply group as well as across groups). In line with the literature, we find that the estimated elasticities are rather robust regarding the specification of the

¹⁰We aggregated semi-flex couples and couples with two flexible partners to calculate elasticities.

utility function, as well as the implementation of observed and unobserved heterogeneity. This is reassuring as it shows that the frequently applied specifications do not restrict the labor supply decision a priori. The only (weak) exception seems to be the implementation of hours restrictions or fixed costs, which tend to drive extensive elasticities up. This finding supports the view that jobs with very few weekly working hours are harder to find than regular part-time or full-time jobs. It is thus more likely that people switch from non-participation to 20 or 40 than to 5 or 10 hours of work when accounting for this fact, which leads to higher elasticities at the extensive margin.

Figure 1: Labor supply elasticities of converged models



Substantially more of the variation can be explained when analyzing the impact of the wage imputation and the handling of wage prediction errors. Our results hold the important message that this part of the model specification is much more relevant to the estimated elasticities than the utility specification. For instance, using predicted wages not only for non-workers but also for the full sample roughly doubles the estimated elasticities when not accounting for the wage prediction error. This substantial difference can be explained by the fact that predicting wages for the full sample reduces the variance of the wage distribution substantially. To explain the observed working hours with less variation in wages and thus income and consumption,

the implied elasticities have to increase. Accounting for wage prediction errors and integrating them out during the estimation markedly reduces the difference. Interestingly, the results differ substantially depending on whether a single random draw or higher numbers are used. The ad hoc procedure of adding a single random draw tends to cancel out the effect of a full sample prediction. In contrast, correcting for the wage prediction error tends to reduce the elasticities, but we still observe the estimated elasticities to be significantly higher than those in which the wage rates were only imputed for non-workers.

Robustness. We performed a wide range of robustness checks to confirm that our results are not special to the used data and methods. In particular, we also used a different wave from the same data set and performed our analysis also using data from the CPS for the US (see table 8 in the online appendix). In addition to the marginal impact (holding all other specification details constant), we investigated the partial impact of the modeling assumptions (in table 9 in the online appendix), which only shows the differences in means due to the specific assumptions (e.g. mean of elasticities using a translog utility specification vs. mean of elasticities using other functional forms, irrespective of other modeling issues). The results we obtained were qualitatively the same. We also checked the robustness regarding the calculation of elasticities and found no differences whether we simulated 1 % or 10 % changes in the own-wage rate (table 10). Also switching the calculation of the elasticities from aggregated to mean, median or other quantile measures did not affect our findings (tables 11 and 12).

Summary. Our results partly confirm previous findings in the literature. While the empirical specification of the systematic utility function has an impact on the statistical fit, we find only little differences in the estimated elasticities. It thus may be justified to rely on simpler model setups when the computational burden is a major concern. However, the majority of applied robustness checks focus on the effects of different utility specifications and usually ignore how the underlying wage treatment may influence the results. We find that these assumptions explain much more variation in outcomes than the specification of the utility function. Most previous robustness checks have thus concentrated on issues of secondary order. Instead, more attention should be paid to the wage imputation and the handling of wage prediction errors. Modeling choices regarding the wage handling may thus also explain part of the large variation found in labor supply studies.

5 Joint estimation of wages and preferences

Our analysis shows that the wage imputation procedure and the handling of the wage prediction error have a huge impact on estimated labor supply elasticities. Nonetheless, it is common practice to estimate the labor supply decision conditional on observed or predicted wages. Wage rates are estimated beforehand and treated as exogenous within the labor supply estimation. This procedure reduces the computational burden, but is naturally rather restrictive. While there are some Hausman-type studies that loosen this fairly strong assumption and find correlation between wages and hours of work (Moffitt, 1984, Tummers and Woittiez, 1991), only little effort has been taken so far in the context of discrete choice labor supply models. Aaberge et al. (1995) and follow-ups estimate labor supply on a random choice set based upon draws from hours and wage distributions they estimate simultaneously. Breunig et al. (2008) and Blundell and Shephard (2012) assume a fixed individual-specific wage rate but allow one specific preference parameter to be correlated with the error term of the wage equation. Although this accounts for at least some interaction between preferences and wages, it still assumes that the labor supply decision is exogenous to the wage rate. Moreover, the correlation structure is rather arbitrary and restrictive. For instance, one may think of potential correlation between the wage rate and several different preference parameters.

This issue is especially important as we expect that correlations between preferences and wages may drive the estimated elasticities. Consider two identical workers with the same observed characteristics but different unobservable productivity and thus wage rates. Suppose in case A that there are indeed no feedback effects between wages and preferences. Subsequently both workers will have the same preferences and their reaction to wage increases will only differ due to their initial consumption level. Now consider case B, in which preferences for, say, consumption are negatively correlated with productivity. Hence, the less productive worker has now a higher preference for consumption, whereas preferences of the more productive worker stay constant. As the less productive worker now values consumption and income more compared to case A, it follows that his labor supply elasticity increases. We show this mechanism in a simple model at the end of Appendix A.

5.1 Estimation setup

We propose a very flexible estimation strategy that overcomes the restrictive exogeneity assumptions of the standard estimation procedure in equation (5) to examine this issue empirically. More specific, we allow the wage rate to depend on hours of work

and preferences for leisure and consumption to be correlated with the error term of the wage equation. However, to render this model feasible, we have to impose some distributional assumptions on the random terms. More precisely, preferences for consumption and leisure and the wage equation residuals are assumed to follow a multivariate normal distribution. We assume that wages are log-normally distributed and estimate them on tenure, education and dummies for having migrated, living in East Germany, being handicapped or working either part-time or overtime. Labor supply and wages are estimated jointly using a full information maximum likelihood framework:

$$\begin{aligned}
\ln(SL) = & \sum_{n \in E} \ln \left(\frac{1}{R} \sum_{r=1}^R P \left\{ U_{ni}^{(r)} > U_{nj}^{(r)}, \forall j \neq i \mid \beta_n^{(r)}, \beta_w, \epsilon_{w,n}^{(r)} = \frac{\ln w_{ni} - \mathbf{x}_{w,ni} \beta_w'}{\sigma_w} \right\} \right) \\
& + \sum_{n \in U} \ln \left(\frac{1}{R} \sum_{r=1}^R P \left\{ U_{ni}^{(r)} > U_{nj}^{(r)}, \forall j \neq i \mid \beta_n^{(r)}, \beta_w, \epsilon_{w,n}^{(r)} \right\} \right) \\
& + \sum_{n \in E} \left(\ln \phi \left\{ \frac{\ln w_{ni} - \mathbf{x}_{w,ni} \beta_w'}{\sigma_w} \right\} - \ln \sigma_w \right) \tag{8}
\end{aligned}$$

where U denotes the subset of unemployed individuals and $\beta_n^{(r)}$, $\epsilon_{w,n}^{(r)}$ are the (randomly drawn) preference coefficients and the error terms of the log-wage equation, respectively. $\phi(\cdot)$ is the density of a standard normal distribution. This framework relates to the estimation approach outlined in Blundell and Shephard (2012), although we approximate the likelihood function using a set of R Halton draws instead of Gaussian Hermite quadrature. Fitting preferences and wages jointly makes it possible to estimate both the influence of hours of work on the wage rate and the variance-covariance matrix of preferences and wages. See Appendix A for details on the model.

Identification. In order to separate and identify both effects properly, we use the actual wage equation residual for workers whose wage rate is observed (subset E), whereas we use multivariate normally distributed Halton sequences to integrate over a set of possible wage equation errors for unemployed individuals. While the standard model assumes preferences and wages to be exogenous, we allow for (multivariate normal) correlation patterns. This makes our estimation approach more flexible, but in turn places another functional form assumption on the estimated covariances.

Chiou and Walker (2007) show that the use of small numbers of draws may mask identification of mixed logit models. We therefore run our estimations using different numbers of draws. While most studies rely on even fewer than 50 (Halton) draws, we use up to 4,000 Halton draws to check the robustness of our results. Reassuringly the estimated variance-covariance matrix is robust to the approximation in terms of both

the coefficients and their standard errors with as few as 50 draws. Table 7 in Appendix B shows the estimation results for our most flexible model using different numbers of draws, varying between 10 and 4,000, indicating that the results are remarkably stable when using at least 50 draws.

5.2 Empirical results

Due to computational constraints, we only estimate our model for the subset of single females in our sample. Table 5 summarizes the results regarding the correlation patterns for this group. We do not present the complete table of estimates here to save space. Full results can be found in Appendix B (table 6).¹¹ The different models are defined as follows:

- (1) Common two-step estimation assuming wages as exogenous.
- (2) Estimates preferences and wages jointly.
- (3) Extends (2) by allowing for hours-dependent wages.
- (4) Extends (3) by additionally allowing for correlation between wages and preferences for consumption.
- (5) Extends (4) by additionally allowing for correlation between wages and taste for leisure.

Our results show that there is indeed correlation between wages and hours of work. For single women, working part-time yields a wage decrease of roughly 7% on average compared to a typical full-time job (see column (5)). The results are statistically significant different from zero at the 5% level. Working 50 or more hours a week is also connected to a decrease in hourly wages compared to full-time employment, although the effect is economically small and statistically not different from zero. These findings indicate an inverted U-shaped relationship between wages and hours of work, and thereby confirm the findings of Moffitt (1984) within the classical continuous hours approach. The estimated variance-covariance matrix between wages and preferences for consumption and leisure shows that there is also significant correlation, whereby we find that higher wages are strongly negatively correlated with preferences for consumption and positively correlated with preferences for leisure (suggesting a backward bending labor supply curve).

¹¹The estimated labor supply and wage equations show the expected signs for the covariates.

Table 5: Joint estimation results single females

	(1)	(2)	(3)	(4)	(5)
Log-wages					
<i>Part time</i>			0.038 (0.030)	-0.041 (0.032)	-0.075** (0.031)
<i>Over time</i>			-0.107*** (0.029)	-0.053 (0.036)	-0.016 (0.029)
<i>N</i>	7455	7455	7455	7455	7455
Pseudo R^2	0.175	0.465	0.466	0.475	0.477
Log-likelihood	-1710	-2228	-2220	-2186	-2176
AIC	3461	4513	4503	4438	4419
BIC	3607	4714	4717	4666	4654
Halton draws	50	50	50	50	50
Correlation					
$\rho_{C,\ln W}$				-0.774*** (0.061)	-0.717*** (0.088)
$\rho_{L,\ln W}$					0.511*** (0.107)
Mean 10% Elasticities					
<i>Intensive</i>	0.210	0.259	0.285	0.456	0.575
<i>Extensive</i>	0.308	0.450	0.457	0.789	1.004
<i>Total</i>	0.222	0.283	0.307	0.501	0.631
Aggregated 10% elasticities					
<i>Intensive</i>	0.172	0.200	0.227	0.365	0.461
<i>Extensive</i>	0.030	0.039	0.041	0.063	0.076
<i>Total</i>	0.201	0.239	0.268	0.428	0.537

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Model (1) corresponds to a common two-step estimation assuming wages as exogeneous. Model (2) estimates preferences and wages jointly. Model (3) allows in addition wages to depend on hours of work. Model (4) also allows for correlation between wages and preferences for consumption. Model (5) extends model (4) by allowing in addition for correlation between wages and taste for leisure. Note that table 7 in Appendix B shows the estimation results for model (5) using different numbers of draws, varying between 10 and 4,000.

The statistical criteria show that models (3) to (5) clearly outperform model (2) irrespective of the specific measure.¹² Even more importantly, both intensive and extensive elasticities rise when the model allows for more flexible preference and wage patterns (compare model (1) to the other four models). In fact, switching from the most restrictive (1) to the most flexible model (5) more than doubles the estimated elasticities from above 0.2 to around 0.6—despite using the same data and identify-

¹²An often used approach to evaluate the performance of different estimators is to run Monte Carlo simulations. However, in our setting, the (joint) significance of the additional regressors and the information criteria presented below allow us to conclude whether our flexible model outperforms the standard approach. In addition, the outcome of the Monte Carlo study depends on the data generating process. If it is chosen with a complex (without a) correlation structure, it would be tautological to conclude that the model allowing for correlation performs best (worst).

ing variation. Indeed, this is strongly in line with the theoretical intuition described above. Our results show that the usual approach to estimate wages separately and independently from the labor supply decision neglects important correlation patterns that determine a households' response to wage changes, driving elasticities towards zero. Hence, when allowing for a more flexible specification, microeconometrically estimated elasticities come closer to macroeconomic ones (Keane, 2011, Chetty et al., 2011, Keane and Rogerson, 2012). One reason for this could be that the more flexible model implicitly captures some of the frictions that are seen as key in explaining these differences. It would be interesting to explicitly incorporate frictions, which are right now captured, among other things, by, e.g., dummies for different hour categories, into the structural estimation and compare the findings to our results in future research.

6 Conclusion

Structural labor supply models are frequently used in the empirical labor supply analysis for many different purposes. In recent years, it has become a standard procedure to estimate labor supply decisions as a choice among a set of different hours alternatives or job opportunities. In contrast to this popularity, little is known about how the numerous modeling assumptions influence the statistical fit, as well as the estimated labor supply elasticities.

In this paper, we provide an overview of the most important specification issues and conduct a comprehensive sensitivity analysis to disentangle the driving factors behind the results from structural labor supply models. Our results show that even if the modeling assumptions concerning the direct utility specification increase or worsen the statistical fit, i.e. the power to explain the observed labor supply behavior, the models are robust in terms of estimated labor supply elasticities. In contrast, the models are highly sensitive to changes in the underlying wage distribution, a mechanisms almost completely neglected in the literature to date. Thus, the question of whether to use predicted or observed wages for actual workers and whether and how to integrate the wage prediction error out during the estimation process has a large and statistically significant impact on the statistical fit of the model and the estimated labor supply elasticities.

We further tackle this issue and propose an alternative estimation method that overcomes the restrictive independence assumptions between wages and work preferences. We allow for correlation between wages and preferences, as well as wage rates that depend on hours of work. Our results show that there are indeed significant cor-

relation patterns in both dimension, which were usually ignored in previous empirical applications. While the standard approach assumes that every worker faces a fixed wage rate irrespective of hours of work, we find that working part-time significantly lowers the hourly wage rate by 7%. Moreover, we find significant correlation between wages and preferences for both leisure and consumption. Our findings clearly reject the exogeneity assumptions that are implicitly made in most discrete choice labor supply applications.

Our findings have important implications for tax policy design. Diamond and Saez (2011) derive simple formulas for the optimal (top) marginal tax rates based upon labor supply elasticities.¹³ They assume an elasticity of 0.25 as an “a mid-range estimate from the empirical literature” which is close to our estimate for the restrictive model. This leads to an optimal top marginal tax rate of $\tau = \frac{1}{1+1.5*0.25} = 72.7\%$. However, an elasticity of 0.6 as found in our most flexible model reduces the optimal tax rate to 52.6% bringing it closer to actually observed values (the top labor tax rate in the US is 42.5%). While we cannot claim that we have identified the true value for the labor supply elasticity—which might not even exist—our analysis shows that more attention should be paid to the specification of structural labor supply models before using them for policy analysis.

¹³The formula for the optimal top tax rate is $\tau = \frac{1-g}{1-g+ae}$ where g is the marginal social welfare weight for the top earners, a is the parameter of the Pareto (income) distribution and e is the labor supply elasticity. Diamond and Saez (2011) assume $g = 0$ to derive the optimal revenue maximizing top tax rate and use an estimated Pareto coefficient of 1.5 for the US.

References

- Aaberge, R., Colombino, U. and Strøm, S. (1999). Labour Supply in Italy: An Empirical Analysis of Joint Household Decisions, with Taxes and Quantity Constraints, *Journal of Applied Econometrics* **14**(4): 403–422.
- Aaberge, R., Colombino, U. and Wennemo, T. (2009). Evaluating Alternative Representations of the Choice Sets in Models of Labor Supply, *Journal of Economic Surveys* **23**(3): 586–612.
- Aaberge, R., Dagsvik, J. K. and Strøm, S. (1995). Labor Supply Responses and Welfare Effects of Tax Reforms, *The Scandinavian Journal of Economics* **97**(4): 635–659.
- Alesina, A. F., Glaeser, E. L. and Sacerdote, B. (2006). Work and Leisure in the United States and Europe: Why So Different?, *NBER Macroeconomics Annual 2005*, Vol. 20, MIT Press, pp. 1–64.
- Bargain, O., Orsini, K. and Peichl, A. (2014). Comparing Labor Supply Elasticities in Europe and the US: New Results, *Journal of Human Resources* (forthcoming).
- Bargain, O. and Peichl, A. (2013). Steady-State Labor Supply Elasticities: An International Comparison, *IZA Discussion Paper* **7698**.
- Blanchard, O. (2007). Comment on “Do Taxes Explain European Employment? Indivisible Labor, Human Capital, Lotteries, and Savings” (by L. Ljungqvist and T. J. Sargent), *NBER Macroeconomics Annual 2006*, Vol. 21, MIT Press, pp. 225–232.
- Bloemen, H. G. and Kapteyn, A. (2008). The estimation of utility-consistent labor supply models by means of simulated scores, *Journal of Applied Econometrics* **23**(4): 395–422.
- Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (1999). Evaluating In-Work Benefit Reform: the Working Families Tax Credit in the UK. paper presented at the Joint Center for Poverty Research conference, Northwestern University.
- Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (2000). The Labour Market Impact of the Working Families’ Tax Credit, *Fiscal Studies* **21**(1): 75–104.
- Blundell, R. and MaCurdy, T. E. (1999). Labor Supply: A Review of Alternative Approaches, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, North-Holland, Amsterdam.
- Blundell, R. and Shephard, A. (2012). Employment, Hours of Work and the Optimal Taxation of Low-Income Families, *Review of Economic Studies* **79**(2): 481–510.

- Borjas, G. J. (1980). The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias, *Journal of Human Resources* **15**(3): 409–423.
- Breunig, R., Cobb-Clark, D. A. and Gong, X. (2008). Improving the Modelling of Couples' Labour Supply, *Economic Record* **84**(267): 466–485.
- Chetty, R. (2012). Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, *Econometrica* **80**(3): 969–1018.
- Chetty, R., Guren, A., Manoli, D. and Weber, A. (2011). Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins, *American Economic Review: Papers and Proceedings* **101**(3): 471–475.
- Chiou, L. and Walker, J. L. (2007). Masking identification of discrete choice models under simulation methods, *Journal of Econometrics* **141**(2): 683–703.
- Dagsvik, J. K., Jia, Z., Orsini, K. and van Camp, G. (2011). Subsidies on low-skilled workers' social security contributions: the case of Belgium, *Empirical Economics* **40**(3): 779–806.
- Dagsvik, J. K. and Strøm, S. (2004). Sectoral Labor Supply, Choice Restrictions and Functional Form, *Discussion Papers Statistics Norway* **388**.
- Dagsvik, J. K. and Strøm, S. (2006). Sectoral Labor Supply, Choice Restrictions and Functional Form, *Journal of Applied Econometrics* **21**(6): 803–826.
- Diamond, P. and Saez, E. (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations, *Journal of Economic Perspectives* **25**(4): 165–190.
- Eissa, N., Kleven, H. J. and Kreiner, C. T. (2008). Evaluation of Four Tax Reforms in The United States: Labor Supply and Welfare Effects For Single Mothers, *Journal of Public Economics* **92**(3-4): 795–816.
- Eklöf, M. and Sacklén, H. (2000). The Hausman-MaCurdy Controversy: Why Do the Results Differ across Studies?, *Journal of Human Resources* **35**(1): 204–220.
- Ericson, P. and Flood, L. (1997). A Monte Carlo Evaluation of Labor Supply Models, *Empirical Economics* **22**(3): 431–460.
- Euwals, R. and van Soest, A. (1999). Desired and actual labour supply of unmarried men and women in the Netherlands, *Labour Economics* **6**(1): 95–118.
- Evers, M., de Mooij, R. and van Vuuren, D. (2008). The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates, *De Economist* **156**(1): 25–43.

- Flood, L., Hansen, J. and Wahlberg, R. (2004). Household Labor Supply and Welfare Participation in Sweden, *Journal of Human Resources* **39**(4): 1008–1032.
- Flood, L., Wahlberg, R. and Pylkkänen, E. (2007). From Welfare to Work: Evaluating a Tax and Benefit Reform Targeted at Single Mothers in Sweden, *LABOUR* **21**(3): 443–471.
- Haan, P. (2006). Much ado about nothing: conditional logit vs. random coefficient models for estimating labour supply elasticities, *Applied Economics Letters* **13**(4): 251–256.
- Hausman, J. A. (1981). Labor Supply, in H. J. Aaron and J. A. Pechman (eds), *How taxes affect economic behavior*, Brookings Institution, pp. 27–72.
- Hoynes, H. W. (1996). Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP, *Econometrica* **64**(2): 295–332.
- Immervoll, H., Kleven, H. J., Kreiner, C. T. and Verdelin, N. (2011). Optimal tax and transfer programs for couples with extensive labor supply responses, *Journal of Public Economics* **95**(11): 1485–1500.
- Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics, *Journal of Econometrics* **156**(1): 3–20.
- Keane, M. P. (2011). Labor Supply and Taxes: A Survey, *Journal of Economic Literature* **49**(4): 961–1075.
- Keane, M. P. and Moffitt, R. (1998). A Structural Model of Multiple Welfare Program Participation and Labor Supply, *International Economic Review* **39**(3): 553–589.
- Keane, M. P. and Wasi, N. (2012). Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior, *Journal of Applied Econometrics* **28**(6): 1018–1045.
- Keane, M. and Rogerson, R. (2012). Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom, *Journal of Economic Literature* **50**(2): 464–476.
- Löffler, M. (2013). Fitting Complex Mixed Logit Models: An Application to Labor Supply. Mimeo, presented at the Stata Conference in New Orleans, July 2013.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.
- MaCurdy, T. E., Green, D. and Paarsch, H. (1990). Assessing Empirical Approaches for Analyzing Taxes and Labor Supply, *Journal of Human Resources* **25**(3): 412–490.

- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.
- Moffitt, R. (1983). An Economic Model of Welfare Stigma, *American Economic Review* **73**(5): 1023–1035.
- Moffitt, R. (1984). The Estimation of a Joint Wage-Hours Labor Supply Model, *Journal of Labor Economics* **2**(4): 550–566.
- Peichl, A., Schneider, H. and Siegloch, S. (2010). Documentation IZAΨMOD: The IZA Policy Simulation Model, *IZA Discussion Papers* **4865**.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*, second edn, Cambridge University Press.
- Tummers, M. P. and Woittiez, I. (1991). A Simultaneous Wage and Labor Supply Model With Hours Restrictions, *Journal of Human Resources* **26**(3): 393–423.
- van Soest, A. (1995). Structural Models of Family Labor Supply – A Discrete Choice Approach, *Journal of Human Resources* **30**(1): 63–88.
- van Soest, A. and Das, M. (2001). Family Labor Supply and Proposed Tax Reforms in the Netherlands, *De Economist* **149**(2): 191–218.
- van Soest, A., Das, M. and Gong, X. (2002). A structural labour supply model with flexible preferences, *Journal of Econometrics* **107**(1-2): 345–374.
- Wagner, G. G., Frick, J. R. and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, *Journal of Applied Social Science Studies* **127**(1): 139–169.
- Ziliak, J. P. and Kniesner, T. J. (1999). Estimating life-cycle labor supply tax effects, *Journal of Political Economy* **107**(2): 326–359.

A Joint estimation

For the joint maximum likelihood estimation in Section 5, we use a Box-Cox transformed utility specification. Thus, the systematic utility of individual n choosing alternative j is given by:

$$v(C_{nj}, L_j) = (\mathbf{x}_n \boldsymbol{\beta}'_1 + \beta_{C,n}) C_{nj}^{(\lambda_C)} + \beta_2 C_{nj}^{(\lambda_C)} L_j^{(\lambda_L)} + (\mathbf{x}_n \boldsymbol{\beta}'_3 + \beta_{L,n}) L_j^{(\lambda_L)} + \beta_{FC,n} 1^{(h_j > 0)} + \beta_4 1^{(h_j = 20)} + \beta_5 1^{(h_j = 40)} \quad (9)$$

with $C_{nj}^{(\lambda_C)}$ and $L_j^{(\lambda_L)}$ defined as:

$$C_{nj}^{(\lambda_C)} = \begin{cases} \frac{C_{nj}^{*\lambda_C} - 1}{\lambda_C} & \text{if } \lambda_C \neq 0 \\ \ln C_{nj}^* & \text{if } \lambda_C = 0 \end{cases} \quad L_j^{(\lambda_L)} = \begin{cases} \frac{L_j^{*\lambda_L} - 1}{\lambda_L} & \text{if } \lambda_L \neq 0 \\ \ln L_j^* & \text{if } \lambda_L = 0 \end{cases} \quad (10)$$

$$C_{nj}^* = C_{nj} / 1000 \quad L_{nj}^* = L_{nj} / 80 \quad (11)$$

and wages given by the wage equation:

$$\ln w_n = \mathbf{z}_n \boldsymbol{\gamma}' + \epsilon_{w,n} \quad (12)$$

We further assume that preferences $(\beta_{C,n}, \beta_{L,n}, \beta_{FC,n})$ and wages (the residual $\epsilon_{w,n}$ of the log-wage equation) are multivariate normally distributed:

$$\begin{pmatrix} \epsilon_{w,n} \\ \beta_{C,n} \\ \beta_{L,n} \\ \beta_{FC,n} \end{pmatrix} \sim N \left(\boldsymbol{\mu} = \begin{bmatrix} 0 \\ \beta_C \\ \beta_L \\ \beta_{FC} \end{bmatrix}, \boldsymbol{\Omega} = \begin{bmatrix} \sigma_w^2 & & & \\ \sigma_{C,w} & \sigma_C^2 & & \\ \sigma_{L,w} & \sigma_{C,L} & \sigma_L^2 & \\ 0 & \sigma_{C,FC} & \sigma_{L,FC} & \sigma_{FC}^2 \end{bmatrix} = \mathbf{L}\mathbf{L}' \right) \quad (13)$$

Instead of estimating the variance-covariance-matrix directly, we estimate its Cholesky decomposition \mathbf{L} . As we cannot directly simulate from the joint distribution, we draw R Halton sequences $\mu_{i,n}^{(r)}$ ($i = 1, \dots, 4$, $r = 1, \dots, R$) from a standard normal and transform them accordingly to the estimated factorization \mathbf{L} :

$$\begin{pmatrix} \hat{\epsilon}_{w,n}^{(r)} \\ \hat{\beta}_{C,n}^{(r)} \\ \hat{\beta}_{L,n}^{(r)} \\ \hat{\beta}_{FC,n}^{(r)} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\beta}_C \\ \hat{\beta}_L \\ \hat{\beta}_{FC} \end{pmatrix} + \begin{pmatrix} \hat{l}_{w,w} & 0 & 0 & 0 \\ \hat{l}_{C,w} & \hat{l}_{C,C} & 0 & 0 \\ \hat{l}_{L,w} & 0 & \hat{l}_{L,L} & 0 \\ 0 & 0 & 0 & \hat{l}_{FC,FC} \end{pmatrix} \begin{pmatrix} \mu_{1,n}^{(r)} \\ \mu_{2,n}^{(r)} \\ \mu_{3,n}^{(r)} \\ \mu_{4,n}^{(r)} \end{pmatrix} \quad (14)$$

We use the actual wage equation residual if the wage is observed, thus $\mu_{1,n}^{(r)} =$

$(\ln w_n - x_{w,n} \hat{\beta}'_w) / \sigma_w$ and $\mu_{i,n}^{(r)} \sim N(0, 1)$ ($i = 2, 3, 4$). If individual n is unemployed and we do not observe her wage rate, we also draw her log-wage equation residual from a normal distribution, $\mu_{1,n}^{(r)} \sim N(0, 1)$.

Using a simplified version of the above model makes it possible to show how the commonly neglected covariance terms influence resulting labor supply elasticities. Consider the utility function outlined in equation (9) and set $\beta_1, \beta_2, \beta_3, \beta_{FC,n}, \beta_4, \beta_5, \lambda_C$ and λ_L equal to zero and give $\beta_{L,n}$ zero variance for simplicity. Now let us assume that leisure is the residual of total time of the week minus hours worked ($L_j = T - H_j$) and consumption is a function of wages, hours and non-labor income ($C_{nj} = w_n H_j + R_n$). The agent's problem can subsequently be written as:

$$\max_H \beta_{C,n} \ln C_{n,j} + \beta_{L,n} \ln L_j = \max_H \beta_{C,n} \ln(w_n H_j + R_n) + \beta_{L,n} \ln(T - H_j). \quad (15)$$

It follows from utility maximization that the labor supply function is given by:

$$H_n^* = \frac{\beta_{C,n} T}{\beta_{L,n} + \beta_{C,n}} - \frac{\beta_{L,n} R_n}{\beta_{L,n} + \beta_{C,n}} w_n^{-1}. \quad (16)$$

We can derive individual labor supply elasticities as:

$$\epsilon_{H,w,n} = \frac{\partial H_n^*}{\partial w_n} \frac{w_n}{H_n^*} = \frac{\beta_{L,n} R_n / T}{\beta_{C,n} w_n - \beta_{L,n} R_n / T}. \quad (17)$$

The elasticity depends inversely on the product of preferences for consumption $\beta_{C,n}$ and hourly wage rates w_n . Moreover, $\beta_{C,n}$ and $\ln w_n$ are assumed to be bivariate normally distributed (see equation (13)). It follows that labor supply elasticities decrease with increasing covariance between wages and consumption preferences.

B Joint estimation results

Table 6: Joint estimation results single females—full results

	(1)	(2)	(3)	(4)	(5)
Preferences					
$C \times \text{Age}$	0.678*** (0.204)	0.920*** (0.315)	0.698*** (0.184)	0.569*** (0.213)	0.591*** (0.221)
$C \times \text{Age}^2$	-0.008*** (0.002)	-0.011*** (0.004)	-0.008*** (0.002)	-0.007*** (0.003)	-0.007*** (0.003)
$C \times \text{Handicap}$	0.527 (1.550)	1.457 (2.044)	1.194 (1.627)	1.507 (1.936)	1.675 (1.909)

<i>C</i>	-8.253** (3.910)	-7.741 (4.746)	-7.740** (3.458)	-1.031 (3.964)	-0.194 (3.936)
<i>C × L</i>	0.801** (0.354)	1.042* (0.542)	0.491 (0.392)	0.797* (0.466)	-0.352 (0.475)
λ_C	0.255 (0.183)	0.164 (0.140)	0.302 (0.202)	0.485*** (0.120)	0.330** (0.130)
<i>L × Age</i>	0.093 (0.095)	0.075 (0.105)	0.104 (0.084)	0.059 (0.116)	0.043 (0.103)
<i>L × Age</i> ²	-0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.000 (0.001)	0.000 (0.001)
<i>L × Handicap</i>	0.440 (0.785)	0.888 (0.922)	0.592 (0.751)	0.809 (1.071)	0.610 (0.938)
<i>L × Care</i>	2.624** (1.284)	3.700* (1.890)	2.749** (1.361)	4.337** (1.938)	4.000** (1.787)
<i>L × Children</i> ₂	1.536* (0.894)	1.847 (1.135)	1.575* (0.892)	2.113* (1.185)	2.054* (1.099)
<i>L × Children</i> _{3,6}	0.456 (0.306)	0.401 (0.358)	0.436 (0.296)	0.220 (0.372)	0.233 (0.333)
<i>L × Children</i> _{7,16}	0.402** (0.160)	0.393** (0.192)	0.384** (0.153)	0.364* (0.213)	0.348* (0.197)
<i>L</i>	-1.296 (1.875)	0.382 (1.867)	-1.204 (1.491)	1.448 (2.317)	2.787 (2.025)
λ_L	-0.775** (0.371)	-0.988*** (0.315)	-0.754*** (0.275)	-0.736** (0.308)	-1.013*** (0.320)
Restrictions					
<i>Fixed costs</i>	-0.745** (0.319)	-1.000*** (0.232)	-1.109*** (0.185)	-1.078*** (0.209)	-1.049*** (0.206)
1(<i>h</i> = 20)	-0.071 (0.123)	-0.063 (0.124)	-0.029 (0.126)	-0.070 (0.126)	-0.124 (0.128)
1(<i>h</i> = 40)	0.811*** (0.089)	0.739*** (0.095)	0.628*** (0.128)	0.411** (0.195)	0.407** (0.201)
Log-wages					
<i>Tenure</i>		0.051*** (0.005)	0.052*** (0.005)	0.049*** (0.004)	0.045*** (0.004)
<i>Tenure</i> ² /100		-0.089*** (0.014)	-0.091*** (0.014)	-0.084*** (0.013)	-0.075*** (0.012)
<i>University</i>		0.295*** (0.033)	0.310*** (0.033)	0.276*** (0.031)	0.253*** (0.029)
<i>Unskilled</i>		-0.259*** (0.057)	-0.266*** (0.057)	-0.272*** (0.051)	-0.255*** (0.047)
<i>East</i>		-0.287*** (0.037)	-0.286*** (0.037)	-0.242*** (0.035)	-0.209*** (0.033)
<i>Foreigner</i>		-0.072 (0.072)	-0.070 (0.071)	-0.060 (0.065)	-0.045 (0.059)
<i>Handicapped</i>		-0.192**	-0.205**	-0.195**	-0.179**

		(0.082)	(0.082)	(0.084)	(0.083)
<i>Part time</i>			0.038	-0.041	-0.075**
			(0.030)	(0.032)	(0.031)
<i>Over time</i>			-0.107***	-0.053	-0.016
			(0.029)	(0.036)	(0.029)
<i>Constant</i>		2.169***	2.170***	2.219***	2.251***
		(0.030)	(0.034)	(0.034)	(0.033)
<hr/>					
Cholesky matrix					
$l_{\ln W}$				0.450***	0.447***
				(0.011)	(0.011)
$l_{C,\ln W}$				-2.249***	-2.040***
				(0.640)	(0.529)
l_C	0.249	2.838**	1.361***	1.838***	1.982***
	(1.158)	(1.127)	(0.504)	(0.682)	(0.723)
$l_{L,\ln W}$					0.957***
					(0.286)
l_L	0.286	1.496***	0.095	1.605**	1.607***
	(0.889)	(0.543)	(0.382)	(0.684)	(0.487)
l_{FC}	0.939	0.842	0.220	0.413	0.357
	(0.730)	(0.614)	(0.477)	(0.570)	(0.575)
<hr/>					
<i>N</i>	7455	7455	7455	7455	7455
Pseudo R^2	0.175	0.465	0.466	0.475	0.477
Log-likelihood	-1710	-2228	-2220	-2186	-2176
AIC	3461	4513	4503	4438	4419
BIC	3607	4714	4717	4666	4654
Halton draws	50	50	50	50	50
<hr/>					
Correlation					
$\rho_{C,\ln W}$				-0.774***	-0.717***
				(0.061)	(0.088)
$\rho_{L,\ln W}$					0.511***
					(0.107)
<hr/>					
Aggregated 10% elasticities					
<i>Intensive</i>	0.172	0.200	0.227	0.365	0.461
<i>Extensive</i>	0.030	0.039	0.041	0.063	0.076
<i>Total</i>	0.201	0.239	0.268	0.428	0.537
Mean 10% Elasticities					
<i>Intensive</i>	0.210	0.259	0.285	0.456	0.575
<i>Extensive</i>	0.308	0.450	0.457	0.789	1.004
<i>Total</i>	0.222	0.283	0.307	0.501	0.631

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Joint estimation results single females—sensitivity of model (5) w.r.t. number of Halton draws

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Preferences							
$C \times Age$	0.463*** (0.168)	0.591*** (0.221)	0.538** (0.211)	0.546** (0.219)	0.543** (0.217)	0.540** (0.215)	0.540** (0.215)
$C \times Age^2$	-0.006*** (0.002)	-0.007*** (0.003)	-0.007*** (0.003)	-0.007*** (0.003)	-0.007*** (0.003)	-0.007*** (0.003)	-0.007*** (0.003)
$C \times Handicap$	1.525 (1.590)	1.675 (1.909)	1.379 (1.846)	1.309 (1.855)	1.250 (1.799)	1.264 (1.808)	1.299 (1.818)
C	-1.943 (3.312)	-0.194 (3.936)	0.458 (3.891)	0.829 (4.023)	0.867 (3.998)	0.683 (3.960)	0.742 (3.969)
$C \times L$	-0.286 (0.495)	-0.352 (0.475)	-0.311 (0.459)	-0.324 (0.467)	-0.313 (0.476)	-0.306 (0.472)	-0.312 (0.474)
λ_C	0.315** (0.148)	0.330** (0.130)	0.349*** (0.135)	0.346*** (0.134)	0.350*** (0.135)	0.351*** (0.136)	0.349*** (0.136)
$L \times Age$	0.060 (0.106)	0.043 (0.103)	0.020 (0.100)	0.016 (0.101)	0.017 (0.101)	0.018 (0.101)	0.018 (0.101)
$L \times Age^2$	-0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
$L \times Handicap$	0.582 (0.962)	0.610 (0.938)	0.451 (0.895)	0.384 (0.869)	0.345 (0.830)	0.357 (0.849)	0.380 (0.855)
$L \times Care$	3.263** (1.423)	4.000** (1.787)	3.769** (1.697)	3.821** (1.733)	3.819** (1.733)	3.795** (1.714)	3.805** (1.716)
$L \times Children_2$	1.657* (0.879)	2.054* (1.099)	2.050* (1.073)	2.088* (1.096)	2.091* (1.095)	2.076* (1.084)	2.079* (1.085)
$L \times Children_{3,6}$	0.159 (0.302)	0.233 (0.333)	0.290 (0.340)	0.319 (0.357)	0.312 (0.354)	0.303 (0.350)	0.304 (0.351)
$L \times Children_{7,16}$	0.267 (0.169)	0.348* (0.197)	0.362* (0.194)	0.360* (0.194)	0.361* (0.193)	0.364* (0.195)	0.362* (0.194)
L	1.397 (2.055)	2.787 (2.025)	3.010 (1.991)	3.201 (2.037)	3.195 (2.026)	3.113 (2.017)	3.137 (2.022)
λ_L	-0.580** (0.232)	-1.013*** (0.320)	-1.067*** (0.335)	-1.116*** (0.333)	-1.115*** (0.339)	-1.090*** (0.338)	-1.091*** (0.335)
Restrictions							
<i>Fixed costs</i>	-0.949*** (0.187)	-1.049*** (0.206)	-1.076*** (0.201)	-1.088*** (0.202)	-1.086*** (0.202)	-1.081*** (0.201)	-1.081*** (0.201)
$1(h = 20)$	-0.094 (0.124)	-0.124 (0.128)	-0.128 (0.128)	-0.136 (0.129)	-0.136 (0.129)	-0.131 (0.128)	-0.132 (0.128)
$1(h = 40)$	0.497*** (0.159)	0.407** (0.201)	0.428** (0.196)	0.422** (0.200)	0.422** (0.200)	0.424** (0.199)	0.423** (0.199)
Log-wages							
<i>Tenure</i>	0.045*** (0.004)	0.045*** (0.004)	0.045*** (0.004)	0.045*** (0.004)	0.045*** (0.004)	0.045*** (0.004)	0.045*** (0.004)
$Tenure^2 / 100$	-0.075***	-0.075***	-0.075***	-0.075***	-0.075***	-0.075***	-0.075***

	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
<i>University</i>	0.253***	0.253***	0.249***	0.248***	0.248***	0.248***	0.248***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
<i>Unskilled</i>	-0.258***	-0.255***	-0.257***	-0.256***	-0.256***	-0.256***	-0.256***
	(0.047)	(0.047)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
<i>East</i>	-0.209***	-0.209***	-0.207***	-0.206***	-0.205***	-0.206***	-0.206***
	(0.034)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
<i>Foreigner</i>	-0.053	-0.045	-0.047	-0.046	-0.046	-0.046	-0.046
	(0.059)	(0.059)	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)
<i>Handicapped</i>	-0.184**	-0.179**	-0.175**	-0.174**	-0.173**	-0.174**	-0.174**
	(0.082)	(0.083)	(0.083)	(0.083)	(0.083)	(0.083)	(0.083)
<i>Part time</i>	-0.069**	-0.075**	-0.076**	-0.079**	-0.079**	-0.078**	-0.078**
	(0.033)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)	(0.032)
<i>Over time</i>	-0.052	-0.016	-0.011	-0.006	-0.005	-0.008	-0.008
	(0.033)	(0.029)	(0.031)	(0.031)	(0.032)	(0.031)	(0.031)
<i>Constant</i>	2.259***	2.251***	2.252***	2.253***	2.253***	2.253***	2.253***
	(0.034)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)

Cholesky matrix

$l_{\ln W}$	0.444***	0.447***	0.447***	0.447***	0.447***	0.447***	0.447***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$l_{C,\ln W}$	-1.244***	-2.040***	-1.959***	-2.050***	-2.057***	-2.021***	-2.023***
	(0.267)	(0.529)	(0.494)	(0.534)	(0.525)	(0.520)	(0.514)
l_C	0.789*	1.982***	1.791***	1.928**	1.925***	1.876***	1.881***
	(0.403)	(0.723)	(0.692)	(0.748)	(0.719)	(0.712)	(0.706)
$l_{L,\ln W}$	0.820***	0.957***	0.942***	0.980***	0.985***	0.965***	0.969***
	(0.239)	(0.286)	(0.283)	(0.295)	(0.295)	(0.289)	(0.290)
l_L	0.555	1.607***	1.532***	1.609***	1.607***	1.572***	1.578***
	(0.487)	(0.487)	(0.459)	(0.495)	(0.487)	(0.475)	(0.476)
l_{FC}	0.284	0.357	0.084	0.051	-0.054	0.005	-0.002
	(0.310)	(0.575)	(0.663)	(0.704)	(0.677)	(0.671)	(0.675)

N	7455	7455	7455	7455	7455	7455	7455
Pseudo R^2	0.476	0.477	0.477	0.477	0.477	0.477	0.477
Log-likelihood	-2179	-2176	-2176	-2176	-2175	-2176	-2176
AIC	4427	4419	4420	4419	4419	4419	4419
BIC	4662	4654	4655	4655	4654	4654	4654
Halton draws	10	50	250	500	1000	2000	4000

Correlation

$\rho_{C,\ln W}$	-0.845***	-0.717***	-0.738***	-0.728***	-0.730***	-0.733***	-0.732***
	(0.108)	(0.088)	(0.095)	(0.092)	(0.090)	(0.090)	(0.090)
$\rho_{L,\ln W}$	0.828***	0.511***	0.524***	0.520***	0.523***	0.523***	0.523***
	(0.231)	(0.107)	(0.103)	(0.099)	(0.098)	(0.101)	(0.100)

Aggregated 10% elasticities

<i>Intensive</i>	0.463	0.461	0.463	0.462	0.463	0.463	0.463
<i>Extensive</i>	0.074	0.076	0.077	0.077	0.077	0.077	0.077

<i>Total</i>	0.537	0.537	0.540	0.539	0.540	0.540	0.540
Mean 10 % Elasticities							
<i>Intensive</i>	0.573	0.575	0.577	0.580	0.578	0.578	0.578
<i>Extensive</i>	0.922	1.004	1.005	1.011	1.012	1.008	1.009
<i>Total</i>	0.618	0.631	0.635	0.638	0.636	0.636	0.636

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C Online appendix

Table 8: Marginal impact of modeling assumptions (CPS)

	Fit	10 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Quadratic	0.640*** (0.062)	0.217 (0.183)	0.207 (0.185)	0.210 (0.185)
Number of Halton draws	-0.015*** (0.002)	0.022** (0.008)	0.023** (0.008)	0.022** (0.008)
Hours restrictions				
Part-time restrictions	-1.855*** (0.089)	0.420** (0.188)	0.397* (0.189)	0.403* (0.190)
Fixed costs	-1.279*** (0.067)	0.192 (0.120)	0.125 (0.104)	0.142 (0.106)
Observed heterogeneity				
in C only	-0.138*** (0.015)	-0.152*** (0.036)	-0.051 (0.031)	-0.078** (0.031)
in L only	-0.258*** (0.026)	-0.066* (0.036)	-0.080* (0.044)	-0.076* (0.042)
in C and L	-0.309*** (0.027)	-0.115** (0.040)	-0.094* (0.044)	-0.097** (0.043)
Unobserved heterogeneity				
in C only	0.067*** (0.013)	-0.118*** (0.037)	-0.113** (0.044)	-0.114** (0.043)
in L only	0.070*** (0.007)	-0.119** (0.053)	-0.121** (0.054)	-0.120** (0.055)
in C and L	0.046*** (0.009)	-0.089** (0.032)	-0.083** (0.036)	-0.084** (0.035)
in C and L (with correl.)	0.021*** (0.007)	-0.063** (0.029)	-0.058** (0.025)	-0.059** (0.026)
Wage imputation				
Full sample, no correction	-0.111* (0.057)	0.912*** (0.284)	0.918*** (0.300)	0.921*** (0.299)
Full sample, 1 random draw	0.025 (0.046)	0.338 (0.244)	0.428* (0.229)	0.413* (0.230)
Non-workers, 1 random draw	0.030 (0.053)	-0.329 (0.362)	-0.237 (0.362)	-0.255 (0.363)
Constant	0.832*** (0.098)	-0.813*** (0.271)	-0.864*** (0.270)	-0.857*** (0.270)
Labor supply types	Yes	Yes	Yes	Yes
Observations	4305	3439	3439	3439
R ²	0.820	0.353	0.340	0.344

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model—the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Partial impact of modeling assumptions (SOEP)

	Fit	10 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Translog	-0.045* (0.024)	-0.125*** (0.021)	-0.035 (0.045)	-0.047 (0.040)
Quadratic	0.135*** (0.013)	0.067* (0.039)	-0.054 (0.053)	-0.037 (0.046)
Box-Cox	-0.093*** (0.017)	0.061 (0.049)	0.094** (0.034)	0.090** (0.034)
Welfare stigma	0.965*** (0.076)	0.051 (0.061)	0.072 (0.047)	0.071 (0.042)
Number of Halton draws	-0.013*** (0.003)	0.008 (0.007)	-0.003 (0.007)	-0.001 (0.007)
Hours restrictions				
None	1.376*** (0.075)	-0.425*** (0.067)	-0.139*** (0.038)	-0.188*** (0.039)
Part-time restrictions	-1.110*** (0.052)	0.145*** (0.041)	0.013 (0.024)	0.035 (0.026)
Fixed costs	-0.244*** (0.034)	0.278*** (0.033)	0.127*** (0.024)	0.153*** (0.023)
Observed heterogeneity				
None	0.398*** (0.063)	-0.002 (0.038)	-0.035* (0.019)	-0.030 (0.021)
in C only	-0.046** (0.017)	-0.070*** (0.015)	0.042*** (0.014)	0.024* (0.013)
in L only	-0.121*** (0.020)	0.067*** (0.015)	0.028* (0.014)	0.035** (0.014)
in C and L	-0.235*** (0.031)	0.004 (0.022)	-0.036*** (0.010)	-0.030*** (0.010)
Unobserved heterogeneity				
None	0.057 (0.040)	0.090 (0.110)	0.125 (0.117)	0.122 (0.117)
in C only	0.029* (0.015)	0.075** (0.036)	0.013 (0.038)	0.023 (0.038)
in L only	0.050 (0.040)	-0.123 (0.110)	-0.032 (0.110)	-0.047 (0.111)
in C and L	-0.035** (0.015)	0.039 (0.039)	0.006 (0.038)	0.011 (0.038)
in C and L (with correl.)	-0.102** (0.039)	-0.128 (0.102)	-0.124 (0.111)	-0.127 (0.110)
Wage imputation				
Full sample imputation	-0.498*** (0.100)	1.248*** (0.288)	1.313*** (0.294)	1.317*** (0.296)
Error integrated out	-0.037 (0.125)	0.267 (0.351)	0.190 (0.359)	0.207 (0.362)
Full sample, no correction	-0.720*** (0.119)	1.921*** (0.145)	2.033*** (0.144)	2.036*** (0.142)
Full sample, error integrated out	-0.334*** (0.081)	1.004*** (0.239)	0.935*** (0.253)	0.960*** (0.254)
Full sample, 1 random draw	0.143 (0.089)	-0.599** (0.237)	-0.554** (0.258)	-0.569** (0.257)
Non-workers, error integrated out	0.269*** (0.094)	-0.544** (0.227)	-0.606** (0.230)	-0.602** (0.231)
N	16,730	13,219	13,219	13,219

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Marginal impact, aggregated 1 % elasticities (SOEP)

	Fit	1 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Quadratic	0.119*** (0.023)	0.132*** (0.028)	0.030 (0.050)	0.043 (0.043)
Box-Cox	-0.020 (0.026)	0.133*** (0.042)	0.087** (0.035)	0.094** (0.035)
Welfare stigma	0.968*** (0.076)	-0.028 (0.084)	0.084 (0.053)	0.071 (0.044)
Number of Halton draws	-0.010*** (0.001)	0.007* (0.004)	-0.002 (0.004)	-0.001 (0.004)
Hours restrictions				
Part-time restrictions	-1.647*** (0.082)	0.390*** (0.071)	0.134*** (0.038)	0.176*** (0.041)
Fixed costs	-1.093*** (0.070)	0.494*** (0.068)	0.217*** (0.039)	0.264*** (0.041)
Observed heterogeneity				
in C only	-0.335*** (0.057)	-0.057 (0.035)	0.060** (0.022)	0.042* (0.022)
in L only	-0.381*** (0.061)	0.032 (0.037)	0.041* (0.022)	0.041* (0.023)
in C and L	-0.475*** (0.070)	-0.002 (0.044)	0.016 (0.020)	0.013 (0.022)
Unobserved heterogeneity				
in C only	0.005 (0.014)	-0.009 (0.024)	-0.054* (0.031)	-0.048 (0.030)
in L only	0.005 (0.013)	-0.085*** (0.024)	-0.032 (0.028)	-0.040 (0.027)
in C and L	-0.041*** (0.013)	-0.036 (0.025)	-0.068** (0.027)	-0.064** (0.026)
in C and L (with correl.)	-0.119*** (0.016)	-0.087** (0.033)	-0.092** (0.033)	-0.093*** (0.033)
Wage imputation				
Full sample, no correction	-0.811*** (0.119)	2.089*** (0.117)	2.245*** (0.088)	2.248*** (0.084)
Full sample, error integrated out	-0.530*** (0.048)	1.427*** (0.103)	1.398*** (0.114)	1.425*** (0.117)
Full sample, 1 random draw	-0.104** (0.049)	0.086 (0.085)	0.102 (0.078)	0.100 (0.079)
Non-workers, error integrated out	0.000 (0.067)	0.054 (0.062)	0.046 (0.035)	0.048 (0.038)
Non-workers, 1 random draw	0.070 (0.056)	-0.157*** (0.056)	-0.220*** (0.028)	-0.214*** (0.032)
Constant	1.004*** (0.121)	-0.930*** (0.104)	-0.730*** (0.084)	-0.770*** (0.087)
Labor supply types	Yes	Yes	Yes	Yes
Observations	16730	13219	13219	13219
R ²	0.854	0.816	0.880	0.889

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 1 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model—the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Marginal impact, mean 10 % elasticities (SOEP)

	Fit	10 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Quadratic	0.119*** (0.023)	0.100*** (0.033)	0.022 (0.047)	0.041 (0.038)
Box-Cox	-0.020 (0.026)	0.101** (0.041)	0.097*** (0.034)	0.096*** (0.033)
Welfare stigma	0.968*** (0.076)	-0.026 (0.063)	0.034 (0.045)	0.026 (0.039)
Number of Halton draws	-0.010*** (0.001)	0.005 (0.004)	-0.003 (0.004)	-0.001 (0.004)
Hours restrictions				
Part-time restrictions	-1.647*** (0.082)	0.353*** (0.089)	0.169*** (0.045)	0.219*** (0.056)
Fixed costs	-1.093*** (0.070)	0.448*** (0.087)	0.254*** (0.044)	0.307*** (0.053)
Observed heterogeneity				
in C only	-0.335*** (0.057)	0.037 (0.036)	0.048* (0.024)	0.046 (0.027)
in L only	-0.381*** (0.061)	0.187*** (0.040)	0.068*** (0.022)	0.101*** (0.027)
in C and L	-0.475*** (0.070)	0.187*** (0.053)	0.036 (0.022)	0.074** (0.029)
Unobserved heterogeneity				
in C only	0.005 (0.014)	0.001 (0.022)	-0.046 (0.031)	-0.036 (0.029)
in L only	0.005 (0.013)	-0.075*** (0.026)	-0.031 (0.028)	-0.041 (0.026)
in C and L	-0.041*** (0.013)	-0.027 (0.025)	-0.059** (0.026)	-0.053** (0.025)
in C and L (with correl.)	-0.119*** (0.016)	-0.083** (0.036)	-0.097*** (0.033)	-0.098*** (0.033)
Wage imputation				
Full sample, no correction	-0.811*** (0.119)	2.130*** (0.106)	2.264*** (0.092)	2.267*** (0.089)
Full sample, error integrated out	-0.530*** (0.048)	1.265*** (0.134)	1.365*** (0.132)	1.364*** (0.140)
Full sample, 1 random draw	-0.104** (0.049)	0.058 (0.049)	0.146 (0.087)	0.122 (0.078)
Non-workers, error integrated out	0.000 (0.067)	0.035 (0.053)	0.062 (0.042)	0.057 (0.043)
Non-workers, 1 random draw	0.070 (0.056)	-0.164*** (0.028)	-0.220*** (0.034)	-0.210*** (0.035)
Constant	1.004*** (0.121)	-0.966*** (0.119)	-0.749*** (0.091)	-0.814*** (0.097)
Labor supply types	Yes	Yes	Yes	Yes
Observations	16730	13219	13219	13219
R ²	0.854	0.820	0.876	0.883

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and taking the mean individual labor supply response. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model—the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Marginal impact, median 10 % elasticities (SOEP)

	Fit	10 % own wage elasticities		
	AIC	Ext.	Int.	Total
Utility function				
Quadratic	0.119*** (0.023)	0.107*** (0.030)	0.079 (0.063)	0.098* (0.048)
Box-Cox	-0.020 (0.026)	0.084** (0.036)	0.042 (0.044)	0.056 (0.040)
Welfare stigma	0.968*** (0.076)	0.001 (0.059)	0.062 (0.068)	0.059 (0.062)
Number of Halton draws	-0.010*** (0.001)	0.005 (0.005)	-0.005 (0.005)	-0.003 (0.005)
Hours restrictions				
Part-time restrictions	-1.647*** (0.082)	0.370*** (0.084)	0.046 (0.069)	0.116* (0.058)
Fixed costs	-1.093*** (0.070)	0.448*** (0.085)	0.112 (0.076)	0.181*** (0.063)
Observed heterogeneity				
in C only	-0.335*** (0.057)	0.046 (0.035)	0.010 (0.023)	0.013 (0.022)
in L only	-0.381*** (0.061)	0.136*** (0.037)	-0.034 (0.022)	-0.001 (0.023)
in C and L	-0.475*** (0.070)	0.130*** (0.044)	-0.103*** (0.025)	-0.068*** (0.024)
Unobserved heterogeneity				
in C only	0.005 (0.014)	-0.020 (0.027)	-0.098** (0.038)	-0.088** (0.036)
in L only	0.005 (0.013)	-0.077*** (0.026)	-0.010 (0.038)	-0.017 (0.036)
in C and L	-0.041*** (0.013)	-0.052** (0.025)	-0.096*** (0.030)	-0.088*** (0.029)
in C and L (with correl.)	-0.119*** (0.016)	-0.107** (0.040)	-0.132*** (0.036)	-0.122*** (0.036)
Wage imputation				
Full sample, no correction	-0.811*** (0.119)	2.169*** (0.098)	2.052*** (0.123)	2.110*** (0.112)
Full sample, error integrated out	-0.530*** (0.048)	1.278*** (0.140)	1.390*** (0.106)	1.397*** (0.103)
Full sample, 1 random draw	-0.104** (0.049)	0.060 (0.053)	0.227 (0.140)	0.199 (0.127)
Non-workers, error integrated out	0.000 (0.067)	0.042 (0.050)	0.013 (0.035)	0.021 (0.032)
Non-workers, 1 random draw	0.070 (0.056)	-0.187*** (0.033)	-0.265*** (0.047)	-0.266*** (0.042)
Constant	1.004*** (0.121)	-0.940*** (0.114)	-0.529*** (0.104)	-0.623*** (0.096)
Labor supply types	Yes	Yes	Yes	Yes
Observations	16730	13219	13219	13219
R ²	0.854	0.832	0.769	0.806

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and taking the median individual labor supply response. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model—the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$