Discrete Choice Labor Supply Models and Wage Exogeneity*

Max Löffler Andreas Peichl Sebastian Siegloch

Work in progress.
This version: October 31, 2013

Abstract. Discrete choice models have become the workhorse in labor supply analyses. Yet, they are often criticized for being a black box due to numerous underlying modeling assumptions. In this paper, we show how these assumptions affect the estimated labor supply elasticities. We find that the estimates are extremely sensitive to the treatment of wages while the empirical specification and the functional form of the utility function are not crucial for the predictions of the model. As a consequence, we propose a very flexible estimation strategy to loosen the commonly made but highly restrictive assumption of independence between wages and the labor supply decision.

JEL Classification: C25, C52, H31, J22
Keywords: household labor supply, elasticity, random utility models

1 Introduction

The use of structural labor supply estimations has become a standard procedure in the empirical analysis of labor supply for both econometricians and policy makers. While 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

*Löffler and Siegloch are affiliated to the Institute for the Study of Labor (IZA) and University of Cologne. Peichl is affiliated to ZEW, University of Mannheim, CESifo and IZA. Corresponding author: Max Löffler, IZA, P.O. Box 7240, 53072 Bonn, Germany, e-mail: loeffler@iza.org.

Hausman approach (see Hausman, 1981)—more recent models make use of the random utility approach and incorporate the labor supply decision as choice among a set of different hours-income combinations (*discrete choice models*). Starting with the works by Aaberge, Dagsvik and Strøm (1995), van Soest (1995) and Hoynes (1996), a wide range of different empirical specifications of this kind of model has been applied. Despite their popularity very little is known about the impact of the various modeling assumptions on the estimated outcomes, i.e., whether certain specifications systematically improve (or worsen) the statistical fit and how the different assumptions may affect the estimated labor supply elasticities.

We aim to fill this gap by examining the robustness of structural labor supply models with regard to their empirical setup. First, we provide a short introduction on modeling assumptions that can or must be made when specifying a discrete choice labor supply model as well as an overview on specifications frequently used in the literature. Second, we set up and estimate 3,456 models each representing a different combination of these modeling assumptions. Based on the estimation results we gather insights how robust the statistical fit of the models and the estimated labor supply elasticities are with respect to the underlying assumptions. In a third step, we introduce a highly flexible estimation strategy to jointly estimate preferences and wages. This approach overcomes the restrictive independence assumptions that are frequently made in the discrete choice context to facilitate the estimation process.

Our analysis extends existing comparative studies in three ways. First, many of the comprehensive surveys in the empirical labor supply literature focus on the principles of alternative estimation strategies (Blundell and MaCurdy, 1999, Blundell et al., 2007) or cross-country comparisons of empirical findings (Meghir and Phillips, 2010, Bargain et al., 201x). Second, although many important studies provide sensitivity checks to show that their results are robust with respect to the utility specification (see, e.g., van Soest, 1995, Euwals and van Soest, 1999, van Soest and Das, 2001, van Soest et al., 2002, Haan, 2006, Aaberge et al., 2009, Pacifico, 2013), they do not take into account all possible combinations of modeling assumption, which is of course reasonable given the different foci of these papers. Instead, these robustness checks usually narrowed down to a small deviation in just one of the several modeling assumptions. Third, unlike meta-analyses such as the one by Evers et al. (2008), who gather estimates from various studies, we estimate various models on the same data using the same control variables. Thus, the outcome variables, i.e., the statistical fit and the estimated labor supply elasticities, have been derived in a controlled environment.

The results of our analysis confirm previous findings in the literature regarding the insensitivity of the models' predictions with respect to the specification of the functional form of the utility function and the inclusion of observed and unobserved preference heterogeneity, hours restrictions and stigma costs of welfare participation. This finding is reassuring with regard to the precision of discrete choice labor supply models in general. However, we also show that the role of wages is crucial for the estimation results—a so far neglected result in the discrete choice literature. Our results reveal that the estimated outcomes are highly sensitive to the wage imputation procedure which is usually neither motivated by economic theory nor subject to robustness checks. In fact, e.g., the choice between predicting wage rates for the full sample or for non-workers only—both procedures are often used in the literature—may increase the estimated labor supply elasticities by up to 100 percent when the applied model does not account for wage prediction errors. We conclude that the attention of sensitivity analyses has been concentrated on more or less irrelevant factors while the main driving forces have been neglected, i.e., the interactions between wages, working hours and preferences. Therefore, we propose a very flexible estimation strategy to overcome commonly made but highly restrictive exogeneity assumptions with regard to the wage rate and the labor supply decision. The estimation results for this estimation approach show that there is indeed substantial correlation between both preferences and wages as well as wages and hours of work. The usual procedure to estimate wages in the first step and assume a fixed wage rate for every individual in the labor supply estimation, ignores these correlation patters.

The remainder of this paper is organized as follows. Section 2 presents the general modeling framework and a short overview on 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. Section 6 concludes.

2 Model and Existing Literature

Structural labor supply estimations build on the assumption of the well-known neoclassical 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:

$$\max_h U(C, L) = \max_h U(f(wh, I), T - h), \tag{1}$$

where leisure L is denoted as difference between total time endowment T and working hours h. Consumption C depends on working hours, the hourly wage rate w, non-labor income I and the tax benefit system f. We assume a static context which implies that consumption equals disposable income as there is no future utility from saving.

Early labor supply models building on the Hausman approach relied on the maximization of the marginal utility over a continuous set of hours of work. This procedure has proven fairly 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 modern countries. Moreover, it has been shown that the estimated models are very sensitive to the underlying wage distribution (Ericson and Flood, 1997, Eklöf and Sacklén, 2000). As the consistent estimation of this kind 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. By comparing different levels of utility one avoids the cumbersome maximization process of Hausman-type models (Aaberge et al., 2009). Flood and Islam (2005) show that continuous hours models can be approximated rather well by these discrete choice models and thus, the discretization itself is barely restrictive. Depending on the labor market, the assumption of a discrete choice between different working hours or job offers may even be more plausible than assuming a continuous choice set (Dagsvik et al., 2013). We focus our analysis on the discrete choice approach as it has become a standard procedure in the labor supply literature.

2.1 General Model

Econometrically, the discrete choice approach boils down to the representation of the labor supply decision in a random utility model. This implies that the true utility of the household can only partly be observed whereas some factors that determine the household's utility are latent at least to the researcher.

$$\begin{aligned} U\left(C_{ni}, L_i | \boldsymbol{x_{ni}}\right) &= v\left(C_{ni}, L_i | \boldsymbol{x_{ni}}\right) + \epsilon_{ni} \\ U_{ni} &= v_{ni} + \epsilon_{ni} \end{aligned} \tag{2}$$

The utility of household n from choosing alternative i is given by U_{ni}, the observed portion is denoted by the *systematic utility function* v_{ni}, ϵ_{ni} denotes an unobserved error term. In the very basic model it is assumed 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. This assumption may seem rather restrictive at first glance. Dagsvik and Strøm (2004, 2006) 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 2.2). The error terms are assumed to be i.i.d. and follow the extreme value type I distribution with the cumulative distribution function $F_\epsilon(x) = \exp(-\exp(-x))$. This distributional assumption leads directly to the representation of the labor supply decision as *conditional logit model* (McFadden, 1974):

$$P(U_{ni} > U_{nj}, \forall j \neq i) = \exp(v_{ni}) / \sum_{s \in J_n} \exp(v_{ns}). \tag{3}$$

In order to estimate the preference coefficients, one has to evaluate the systematic utility function v for every household $n = 1, \ldots, N$ and every choice category within the choice set J_n. Given the different utility levels, the model can be estimated via maximum likelihood. The derivation of the (log)-likelihood function is very straightforward in this case. However, there are some modeling assumptions that have to be made as well as several possible extensions to this simple setup.

Choice set The first decision in the estimation regards 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 sensitivity 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 and thus 49 alternatives.

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. Stern (1986) discussed the implications of different utility specifications in the context of continuous labor supply models.

2.2 Model extensions

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

Including also unobserved heterogeneity overcomes the IIA assumption as it allows for unobservable variation in preferences between choice alternatives. There are mainly two 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 in this field.

Welfare stigma and benefit take-up Several extensions to the standard model have been proposed in the literature. While the model as described so far assumes that households build their preferences only 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 tied 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) argued that working part-time could be connected with an unobservable disutility as well, because part-time jobs may exhibit higher search costs. Euwals and van Soest (1999) extended this idea and introduced fixed costs of work which have been used in several applications since then. While both approaches help explaining the observed labor market outcomes, their rational remains rather ad hoc. Aaberge et al. (1995) provide a more convenient theoretical framework which delivers a structural interpretation of fixed costs and the utility connected to certain hours alternatives. In their model, households choose between (latent) job offers which differ not only with regard to the working hours but also in terms of wages, non-monetary attributes and availability.

2.3 Wage imputation procedure

In addition to the specification of the utility function, there are important modeling assumptions with regard to 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), 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 the wage rates are not observed or for the full sample (see MaCurdy et al., 1990, for a discussion of the pros and cons 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 or not. Especially when using predicted wages for the full sample, the "new" distribution of wages will typically have a significantly lower variance and at least for some workers, the predicted wage will differ considerably from the observed one. Thus, ignoring the error when predicting wage rates 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). A rather rough approximation that has been used in some applications is to add just a single random draw to the predicted wage rates (Bargain et al., 201x). While this procedure lacks for a theoretical rational, it reduces the computational burden of the estimation substantially.

2.4 Estimation approach

The named extensions especially with regard to 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:

$$L = \prod_{n=1}^{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\exp\left(v_{ni}\left\{\hat{w}_n, \boldsymbol{\beta_u}\right\}\right) g(h_i)}{\sum_{j \in J_n} \exp\left(v_{nj}\left\{\hat{w}_n, \boldsymbol{\beta_u}\right\}\right) g(h_j)} f(\boldsymbol{\beta_u}) f(\hat{w}_n)\, \mathrm{d}\boldsymbol{\beta_u} \mathrm{d}\hat{w}_n, \tag{4}$$

where $i \in J_n$ denotes the alternative chosen by individual n. The likelihood contributions depend not only on the systematical utility function but also on the availability of the choice alternatives which is denoted by $g(h_i)$. This setup implies that the availability of choice alternatives can be separated from the systematic utility which is a reasonable assumption at least for highly regulated labor markets like in many industrialized countries. As the preferences may also include unobserved heterogeneity, the probability that household n maximizes her utility at choice alternative i has to be integrated over the possible set of coefficients $\boldsymbol{\beta_u}$. Similarly, the individual likelihood contributions have to be integrated over the range of possible wage predictions $\hat{w}_n$. As both variables will typically not be uniformly distributed, the choice probability has to be weighted by the (joint) probability density functions of the random components.

The model as written down in equation (4) 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 $r = 1, \ldots, R$ random draws from the distributions of $\boldsymbol{\beta_u}$ and $\hat{w}_n$. The simulated log-likelihood is then given by:

$$\ln(SL) = \sum_{n=1}^{N} \ln \left(\frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left(v_{ni}\left\{\hat{w}_n^r, \boldsymbol{\beta_u^r}\right\}\right) g(h_i)}{\sum_{j \in J_n} \exp\left(v_{nj}\left\{\hat{w}_n^r, \boldsymbol{\beta_u^r}\right\}\right) g(h_j)} \right) \tag{5}$$

When the number of draws goes to infinity, the simulated log-likelihood in (5) converges to the log-likelihood of the model denoted in (4). 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). Details on the estimation procedure can be found in Löffler (2013).

The representation of the labor supply decision as random utility model instead of the traditional continuous hours approach allows us to estimate the model without imposing restrictive a priori assumptions on the coefficients or the functional form of the utility function. Whether or not the estimated coefficients are in line with the economic intuition and thus utility maximizing behavior can be checked afterwards. Euwals and van Soest (1999) point out that it is only necessary to check the marginal utility of consumption as, e.g., fixed costs of work as well as differences between desired and actual hours might explain that for some individuals the marginal utility from leisure is negative. However, some authors penalize the likelihood function to ensure that the utility increases with income for most individuals (Euwals and van Soest, 1999, Bargain et al., 201x).

2.5 Existing Literature

Tables 1 and 2 provide an overview on 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 in the applied works, i.e., either a translog, a quadratic or a Box-Cox transformed specification. As the Stone-Geary function can be interpreted

Table 1: Different model specifications

Paper	Utility Function	Heterogeneity[*] Observed	Heterogeneity[*] Unobs.	Welfare 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[†]	—	—/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. (201x)	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, 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. (201x)	Two step	Yes	Full sample	Random draw

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

as a simplification of the translog as well as 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 less studies allow for preference heterogeneity with regard to 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 at few hours categories, most authors include either fixed costs of working or hours restrictions or both in their models. 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 because it is a common finding that the benefit participation rate deviates substantially from full take-up. Thus, models which 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 joint estimations 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 when the wage rate is unobserved only or for the full sample in order to avoid two distinct wage distributions. With regard to the handling of the wage prediction errors, it becomes more and more common practice to incorporate and integrate the errors out during the estimation.

3 Data

The estimations in this paper are performed on the German Socio-Economic Panel (SOEP), a representative household panel survey for Germany (Wagner et al., 2007). Established in 1984, SOEP was frequently extended with specific subsamples to increase the representativeness for specific subgroups like immigrants or high-income earners. SOEP includes now more than 24,000 individuals in around 11,000 households. We use the 2008 wave of SOEP, which includes household data from the year 2008 as well as data on the labor supply behavior and incomes from the preceding year. We rely on the tax and transfer system of 2007 as well and down rate the household data accordingly. We focus our analysis on the working age population and thus exclude 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 has access to the regular labor market and has a flexible labor supply. Therefore, we exclude households where all decision makers are self-employed, civil servants or in the military service. Moreover, our subsample includes some households with more than two adults, mainly adult children living with their parents. We exclude these individuals from the estimation as it is unclear how their consumption and utility are determined (Dagsvik et al., 2011). The parental household is included in the subsample for our labor supply estimations though.

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 single men and single women either in a single household or living with 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 where 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 where 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 where both partners are flexible regarding their labor supply behavior.

For the computation of consumption levels for the different choice categories, we rely on IZAΨMOD v3.0.0, the policy simulation model of the Institute for the Study of Labor. IZAΨMOD 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 to apply the tax and benefit system for every possible wage rate for every household in every step of the numerical likelihood maximization. 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 m_{nj} into disposable income while controlling for a rich set of household characteristics $\boldsymbol{x_{nj}}$ as well as all available sources of non-labor incomes $\boldsymbol{z_{nj}}$ ($n = 1, \dots, N$, $j \in J_n$):

$$C_{nj} = m_{nj}\beta_{w,1} + m_{nj}^2\beta_{w,2} + \boldsymbol{x_{nj}\beta'_{w,3}}m_{nj} + \boldsymbol{x_{nj}\beta'_{w,4}}m_{nj}^2 + \boldsymbol{z_{nj}B_{x,z}x'_{nj}} + \beta_0 + \eta_{nj}. \quad (6)$$

The resulting R^2 shows a 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 allow for unobserved tax determinants as well, 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 tax and transfer system, we are thus confident that the approximation does not affect our findings.

4 Sensitivity analysis

Although there have been some robustness checks in the literature (see tables 1 and 2), these checks usually narrowed down to a small deviation in just one of the modeling assumptions. On the other contrary, Evers et al. (2008) performed a broader meta-analysis of labor supply models comparing estimated labor supply elasticities for different countries and explained them mainly by study characteristics. In either case, it is hard to draw 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 an controlled environment that is using the same data basis. The estimation results allow us to determine how sensitive or robust the estimated outcomes are with respect to the specification and the wage imputation procedure of the model.

4.1 Analysis setup

To perform our sensitivity-analysis, we combine frequently used modeling assumptions and estimate all possible combinations of these specifications. We estimated 3,456 different model specifications for the five distinct labor supply types which leads us to a total of 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 to estimate this large number of models. Therefore, we decided to drop those estimation results from our analysis that did not converge. Depending on the labor supply group we loose up to 6 percent of the estimation results and end up with a sample of 16,730 different maximum likelihood estimations.

Table 3 shows the different specifications and the number of converged estimation results. We estimated 1,152 distinct models with a Box-Cox transformed utility specification for each of the five labor supply groups. But only 1,022 estimation results for single males and 1,132 for single females are included in our sample. Of all estimated

Table 3: Estimated model combinations

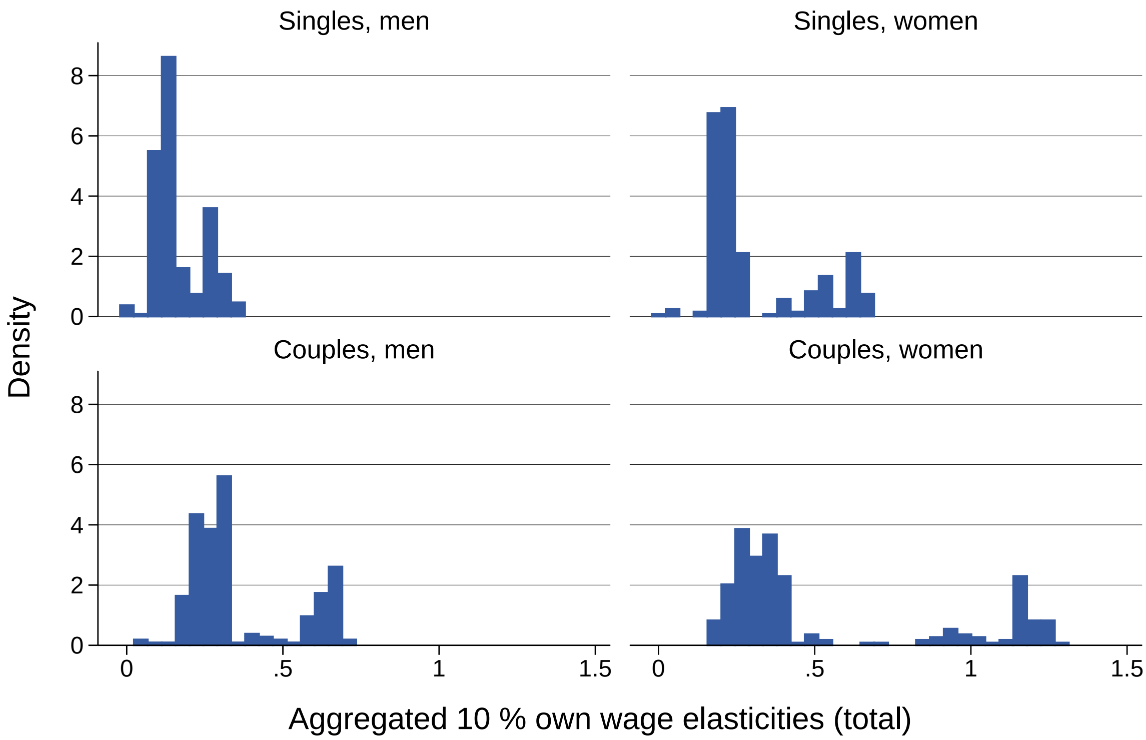
Model Parameter	Option	N	Number of Converged Models* SgM	SgF	CoM	CoF	CoMF
Utility function	Box-Cox	1,152	1,022	1,132	951	1,148	1,029
	Quadratic	1,152	1,152	1,151	1,152	1,133	1,152
	Translog	1,152	1,125	1,144	1,148	1,148	1,143
Welfare stigma	No	1,728	1,642	1,701	1,607	1,713	1,664
	Yes	1,728	1,657	1,726	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.

models (regardless of the functional form of the utility function), 1,152 models neglected any kind of hours restrictions or fixed costs, 1,152 models included part-time restrictions, 1,152 models accounted for fixed costs of work.

Figure 1 shows the distribution of labor supply elasticities across the converged models for the four labor supply types. The graph shows considerable variation across the different modeling setups. In order to make the estimation results comparable, we standardize the statistical fit and the estimated elasticities within a labor supply group. We then run meta-regressions of the estimation results on the different modeling assumptions (mainly represented as dummy variables). We measure the statistical fit by the Akaike and the Bayesian Information Criteria 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.

Figure 1: Labor supply elasticities of converged models



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 worsens the statistical fit by roughly 12 % of a standard deviation in the sample. These results have to be compared to a rather simple reference model using a translog utility 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 lots of statistically significant relationships. However, the presented standard errors are not bootstrapped as this would make 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 0.5 or even 1.0 in absolute values are also economically interesting. In addition to the marginal impact, we present the partial impact of the modeling assumptions in table 5.

Table 4: Marginal impact of modeling assumptions (SOEP)

	Statistical fit		10 % own wage elasticities		
	AIC	BIC	Ext.	Int.	Total
Utility function					
Quadratic	0.119***	0.119***	0.124***	-0.015	0.004
	(0.023)	(0.023)	(0.028)	(0.062)	(0.053)
Box-Cox	-0.020	-0.020	0.116***	0.080**	0.085**
	(0.026)	(0.026)	(0.040)	(0.035)	(0.034)
Welfare stigma	0.968***	0.968***	0.045	0.065	0.065
	(0.076)	(0.076)	(0.062)	(0.047)	(0.042)
Number of Halton draws	-0.010***	-0.010***	0.005	-0.003	-0.002
	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)
Hours restrictions					
Part-time restrictions	-1.647***	-1.647***	0.384***	0.105**	0.152***
	(0.082)	(0.082)	(0.070)	(0.039)	(0.042)
Fixed costs	-1.093***	-1.093***	0.481***	0.187***	0.238***
	(0.070)	(0.070)	(0.067)	(0.040)	(0.041)
Observed heterogeneity					
in C only	-0.335***	-0.335***	-0.049	0.060**	0.043*
	(0.057)	(0.057)	(0.035)	(0.022)	(0.023)
in L only	-0.381***	-0.381***	0.048	0.045**	0.046*
	(0.061)	(0.061)	(0.038)	(0.021)	(0.023)
in C and L	-0.475***	-0.474***	0.016	0.012	0.013
	(0.070)	(0.070)	(0.044)	(0.019)	(0.022)
Unobserved heterogeneity					
in C only	0.005	0.005	-0.006	-0.059*	-0.051
	(0.014)	(0.014)	(0.023)	(0.032)	(0.030)
in L only	0.005	0.005	-0.081***	-0.029	-0.037
	(0.013)	(0.013)	(0.023)	(0.028)	(0.027)
in C and L	-0.041***	-0.041***	-0.037	-0.069**	-0.064**
	(0.013)	(0.013)	(0.024)	(0.027)	(0.026)
in C and L (with correl.)	-0.119***	-0.119***	-0.082**	-0.102***	-0.101***
	(0.016)	(0.016)	(0.034)	(0.034)	(0.033)
Wage imputation					
Full sample, no correction	-0.811***	-0.811***	2.121***	2.235***	2.240***
	(0.119)	(0.119)	(0.094)	(0.091)	(0.086)
Full sample, error integrated out	-0.530***	-0.530***	1.399***	1.385***	1.406***
	(0.048)	(0.048)	(0.119)	(0.123)	(0.124)
Full sample, 1 random draw	-0.104**	-0.104**	0.071	0.131	0.121
	(0.049)	(0.049)	(0.062)	(0.093)	(0.088)
Non-workers, error integrated out	0.000	0.000	0.048	0.040	0.041
	(0.067)	(0.067)	(0.063)	(0.041)	(0.041)
Non-workers, 1 random draw	0.070	0.070	-0.230***	-0.232***	-0.235***
	(0.056)	(0.056)	(0.038)	(0.035)	(0.037)
Constant	1.004***	1.004***	-0.939***	-0.678***	-0.726***
	(0.121)	(0.121)	(0.094)	(0.087)	(0.087)
Labor supply types	Yes	Yes	Yes	Yes	Yes
Observations	16730	16730	13219	13219	13219
R^2	0.854	0.854	0.849	0.870	0.881

Notes: Standard errors clusted by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Partial impact of modeling assumptions (SOEP)

	Statistical fit		10 % own wage elasticities		
	AIC	BIC	Ext.	Int.	Total
Utility function					
Translog	-0.045*	-0.045*	-0.125***	-0.035	-0.047
	(0.024)	(0.024)	(0.021)	(0.045)	(0.040)
Quadratic	0.135***	0.135***	0.067*	-0.054	-0.037
	(0.013)	(0.013)	(0.039)	(0.053)	(0.046)
Box-Cox	-0.093***	-0.093***	0.061	0.094**	0.090**
	(0.017)	(0.017)	(0.049)	(0.034)	(0.034)
Welfare stigma	0.965***	0.965***	0.051	0.072	0.071
	(0.076)	(0.076)	(0.061)	(0.047)	(0.042)
Number of Halton draws	-0.013***	-0.013***	0.008	-0.003	-0.001
	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)
Hours restrictions					
None	1.376***	1.375***	-0.425***	-0.139***	-0.188***
	(0.075)	(0.075)	(0.067)	(0.038)	(0.039)
Part-time restrictions	-1.110***	-1.110***	0.145***	0.013	0.035
	(0.052)	(0.052)	(0.041)	(0.024)	(0.026)
Fixed costs	-0.244***	-0.243***	0.278***	0.127***	0.153***
	(0.034)	(0.034)	(0.033)	(0.024)	(0.023)
Observed heterogeneity					
None	0.398***	0.398***	-0.002	-0.035*	-0.030
	(0.063)	(0.063)	(0.038)	(0.019)	(0.021)
in C only	-0.046**	-0.046**	-0.070***	0.042***	0.024*
	(0.017)	(0.017)	(0.015)	(0.014)	(0.013)
in L only	-0.121***	-0.121***	0.067***	0.028*	0.035**
	(0.020)	(0.020)	(0.015)	(0.014)	(0.014)
in C and L	-0.235***	-0.235***	0.004	-0.036***	-0.030***
	(0.031)	(0.031)	(0.022)	(0.010)	(0.010)
Unobserved heterogeneity					
None	0.057	0.057	0.090	0.125	0.122
	(0.040)	(0.040)	(0.110)	(0.117)	(0.117)
in C only	0.029*	0.029*	0.075**	0.013	0.023
	(0.015)	(0.015)	(0.036)	(0.038)	(0.038)
in L only	0.050	0.050	-0.123	-0.032	-0.047
	(0.040)	(0.040)	(0.110)	(0.110)	(0.111)
in C and L	-0.035**	-0.035**	0.039	0.006	0.011
	(0.015)	(0.015)	(0.039)	(0.038)	(0.038)
in C and L (with correl.)	-0.102**	-0.102**	-0.128	-0.124	-0.127
	(0.039)	(0.039)	(0.102)	(0.111)	(0.110)
Wage imputation					
Full sample imputation	-0.498***	-0.498***	1.248***	1.313***	1.317***
	(0.100)	(0.100)	(0.288)	(0.294)	(0.296)
Error integrated out	-0.037	-0.037	0.267	0.190	0.207
	(0.125)	(0.125)	(0.351)	(0.359)	(0.362)
Full sample, no correction	-0.720***	-0.720***	1.921***	2.033***	2.036***
	(0.119)	(0.119)	(0.145)	(0.144)	(0.142)
Full sample, error integrated out	-0.334***	-0.334***	1.004***	0.935***	0.960***
	(0.081)	(0.081)	(0.239)	(0.253)	(0.254)
Full sample, 1 random draw	0.143	0.143	-0.599**	-0.554**	-0.569**
	(0.089)	(0.089)	(0.237)	(0.258)	(0.257)
Non-workers, error integrated out	0.269***	0.269***	-0.544**	-0.606**	-0.602**
	(0.094)	(0.094)	(0.227)	(0.230)	(0.231)
N	16,730	16,730	13,219	13,219	13,219

Notes: Standard errors clusted by labor supply group and wage imputation procedure. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Goodness of fit Although the statistical fit is usually not the outcome of highest interest, our results show several interesting patterns for future applications. First of all, 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. The performance of the random coefficients models that also allow for unobserved heterogeneity is surprisingly bad compared to the computational burden of their estimation. The results with regard to 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 for the full sample increases the fit significantly. However, this is not surprising as it demonstrates how much of the variation in the data is lost only 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 except for the implementation of fixed costs or hours restrictions there is hardly a single modeling assumption that guarantees a good fit. Instead, there are several small issues that 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. In line with the literature, we find that the estimated elasticities are rather robust with regard to the specification of the 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 jobs with, e.g., 20 hours of work.

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 way more relevant to the estimated elasticities than the utility specification. E.g., using predicted wages not only for non-workers but 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. To account for wage prediction errors and to integrate these errors out during the estimation reduces the difference markedly. Interestingly the results differ a lot 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 the effect of a full sample prediction out. 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 where the wage rates were imputed only 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 Current Population Survey for the US. The results we obtained were qualitatively the same. We also checked the robustness with regard to the calculation of elasticities and found no differences whether we simulated 1 % or 10 % changes in the own-wage rate. Also switching the calculation of the elasticities from aggregated to mean, median or other quantile measures did not affect our findings.

Summary In part our results 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 was focused on the effects of different utility specifications and has usually ignored how the underlying wage treatment may influence the results. We find that these assumptions explain a lot more variation in outcomes than the specification of the utility function. Most previous robustness checks have thus concentrated on rather irrelevant issues. Instead, more attention should be paid to the wage imputation and the handling of wage prediction errors.

5 Joint estimation of wages and preferences

Our analysis shows that the utility specification hardly affects the estimated labor supply elasticities. In contrast, the wage imputation procedure and the handling of the wage prediction error have a huge impact. Despite this importance, it is common practice to estimate the labor supply decision conditional on observed or predicted wages. The wage rates are then estimated beforehand and treated as exogenous within the labor supply estimation. This procedure reduces the computational burden, but is obviously 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 on draws from the hours and wage distribution. 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 restrictive as one may think of potential correlation between the wage rate and disutility components like fixed costs or welfare stigma.

We propose a very flexible estimation strategy that overcomes the restrictive exogeneity assumptions of the standard estimation procedure. More specific, we allow the wage rate to depend on hours of work and preferences for leisure and consumption as well as fixed costs to be correlated with the error term of the wage equation. To make this model feasible, we however 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 estimate log-wages on tenure, labor market experience, education and dummies for foreigners, living in East Germany, being handicapped or working either part-time or overtime. Labor supply and wages are estimated using a full information maximum likelihood framework:

$$\begin{aligned}
\ln(SL) = & \sum_{n\in E} \ln\left(\frac{1}{N}\sum_{r=1}^{R} P(U^r_{ni} > U^r_{nj}, \forall j \neq i | \hat{w}_{nj}, \epsilon^r_{\beta,n}, \epsilon^r_{w,n} = \ln w_{ni} - \boldsymbol{x_{w,ni}}\beta'_w)\right) \\
& + \sum_{n\in U} \ln\left(\frac{1}{N}\sum_{r=1}^{R} P(U^r_{ni} > U^r_{nj}, \forall j \neq i | \hat{w}_{nj}, \epsilon^r_{\beta,n}, \epsilon^r_{w,n})\right) \\
& + \sum_{n\in E} \left(\ln\phi\left\{\frac{\ln w_{ni} - \boldsymbol{x_{w,ni}\beta'_w}}{\sigma_w}\right\} - \ln\sigma_w\right) \qquad (7)
\end{aligned}$$

where U denotes the subset of unemployed individuals and $\epsilon^r_{\beta,n}$, $\epsilon^r_{w,n}$ are the (randomly drawn) error terms from the distribution of random preference coefficients $\boldsymbol{\beta_u}$ and the wage equation, respectively. This framework makes it possible to estimate the influence of hours of work on the wage rate as well as the variance-covariance matrix between preferences and wages. 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.

We estimate our model for single male households. Table 6 summarizes the relevant results with regard to the correlation patterns. Our results show that there is indeed correlation between wages and hours of work (models (3) to (5) in the upper part of table 6), which is well in line with earlier findings. For single men working part-time, i.e., between ten and thirty hours a week, leads to a wage decrease of 14-32 % compared to a typical 40 hours per week employment. The results are statistically highly significant. Working 50 or more hours a week is also connected to a (smaller) decrease in wages compared to full-time employment. This effect varies between 11 and 16 %. These findings indicates 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.

Table 6: Estimation results single males

	(1)	(2)	(3)	(4)	(5)
$\ln w$					
part time			-0.151***	-0.273***	-0.390***
			(0.0511)	(0.0524)	(0.0628)
over time			-0.186***	-0.129***	-0.124***
			(0.0350)	(0.0313)	(0.0370)
$l_{FC,\ln w}$				0.427***	0.0203
				(0.0125)	(0.153)
$l_{C,\ln w}$					-0.0806
					(0.205)
$l_{L,\ln w}$					-0.408***
					(0.0364)
N	5453	5453	5453	5453	5453
r2_p	0.227	0.515	0.520	0.533	0.537
ll	-1172.5	-1537.3	-1521.9	-1480.1	-1467.4
aic	2373.0	3126.7	3099.8	3022.3	3004.7
bic	2465.5	3298.4	3284.7	3227.0	3235.9

Notes: Estimation results using 5 Halton sequences. Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The estimated variance-covariance matrix between wages and preferences for consumption and leisure shows that there is some significant correlation as well (see models (4) and (5) in the lower part of table 6). Here we present cross entries of the Cholesky matrix, showing the correlation patterns between wages and fixed costs of work in model (4) and additionally preferences for consumption and leisure in column (5). We see that there is indeed a significant relationship, which is mainly due to correlation between wages and preferences for leisure.

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 choice among a set of different hours alternatives or job opportunities instead of maximizing the marginal utility over a continuous set of working hours. In contrast to this popularity, little is known on how the numerous modeling assumptions usually made when applying this kind of model impact the statistical fit as well as the estimated outcomes in terms of labor supply reactions to changes in earnings.

In this paper, we provide an overview on the most important specification issues and carry out a comprehensive sensitivity-analysis in order to disentangle the driving factors behind modern 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 when it comes to estimated labor supply elasticities. These results are well in line with several robustness checks that have been applied in the literature. In contrast to the robustness regarding the utility function, the models are highly responsive to changes in the underlying wage distribution. This concerns the wage imputation procedure for non-workers and the decision whether to use observed or predicted wages for actual workers as well as the inclusion of potential wage prediction errors. In fact, our results indicate that, e.g., using predicted wages for the full sample instead of predicting wages for non-workers only, roughly doubles the estimated elasticities when the model does not account for errors in the wage prediction. Thus, 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. While it is common practice in many empirical studies to check the robustness of the results regarding the specification of the direct utility function, surprisingly little effort has been taken so far to check the robustness of the models with respect to the underlying wage treatment.

Therefore, we further tackle this issue and an alternative estimation method that overcomes the restrictive independence assumptions previously made in the context of discrete choice models. We allow for both correlation between wages and preferences, and wage rates that depend on hours of work. Our results show that there is indeed a significant relationship in both directions which is usually ignored in 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 14-32 %. A similar (but smaller) effect is found for working overtime. Moreover we find partly significant correlation between preferences and wages. Our findings clearly reject the exogeneity assumptions that are implicitly made in most discrete choice labor supply applications.

References

Aaberge, R., Colombino, U. and Strøm, S. (1999). Labour supply in italy: An empirical analysis of joint houshold 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.

Bargain, O., Orsini, K. and Peichl, A. (201x). Comparing labor supply elasticities in europe and the us: New results, *Journal of Human Resources* . forthcoming.

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., MaCurdy, T. and Meghir, C. (2007). Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics, *in* J. J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 6a, Elsevier, chapter 69, pp. 4667–4775.

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.

Dagsvik, J. K., Jia, Z., Kornstadt, T. and Thoresen, T. O. (2013). Theoretical and practical arguments for modeling labor supply as a choice among latent jobs, *Journal of Economic Surveys* . doi 10.1111/joes.12003.

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.

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. and Islam, N. (2005). A monte carlo evaluation of discrete choice labour supply models, *Applied Economics Letters* **12**(5): 263–266.

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.

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.

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.

Meghir, C. and Phillips, D. (2010). Labour supply and taxes, *in* J. Mirrless, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds), *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press, pp. 202–274.

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.

Pacifico, D. (2013). On the role of unobserved preference heterogeneity in discrete choice models of labour supply, *Empirical Economics* **45**(2): 929–963.

Peichl, A., Schneider, H. and Siegloch, S. (2010). Documentation IZAΨMOD: The IZA policy simulation model, *IZA Discussion Papers* **4865**.

Stern, N. (1986). On the specification of labour supply functions, *in* R. Blundell and I. Walker (eds), *Unemployment, search and labour supply*, Cambridge University Press, chapter 9, pp. 143–189.

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.