

Löffler, Max; Peichl, Andreas; Sieglöcher, Sebastian

## Conference Paper

# Validating Structural Labor Supply Models

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Labor Supply, No. D17-V2

### Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Löffler, Max; Peichl, Andreas; Sieglöcher, Sebastian (2013) : Validating Structural Labor Supply Models, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Labor Supply, No. D17-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79819>

### 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.*

# Validating Structural Labor Supply Models

Max Löffler\*    Andreas Peichl†    Sebastian Siegloch‡

Work in progress.

This version: February 27, 2013

**Abstract.** Although discrete hours choice models have become the workhorse in labor supply analyses. Yet, they are often criticized for being a black box due to their numerous underlying modeling assumptions, with respect to, e.g., the functional form, unobserved error components or several exogeneity assumptions. In this paper, we open the black box and show how these assumptions affect the statistical fit of the models and, more importantly, the estimated outcomes, i.e., estimated labor supply elasticities. In total we estimate 2,219 different model specifications. Our results show that the specification of the utility function is not crucial for performance and predictions of the model. We find however that the estimates are extremely sensitive to the treatment of the wages—a neglected dimension so far. We show that, e.g., the choice between predicting wages for the full sample instead of using predicted wages only for non-workers—two methods frequently used—increases labor supply elasticities by up to 100 percent. As a consequence, we propose a new estimation strategy which overcomes the highly restrictive but commonly made assumption of independence between wages and the labor supply decision.

**JEL Classification:** C25, C52, H31, J22

**Keywords:** labor supply, elasticity, random utility models, meta-analysis, SOEP

## 1 Introduction

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

---

\*Institute for the Study of Labor (IZA) and University of Cologne (loeffler@iza.org).

†IZA, University of Cologne, ISER and CESifo (peichl@iza.org).

‡IZA and University of Cologne (sieglloch@iza.org).

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)—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 in the last two decades. Despite their popularity very little is known about the impact of the various modeling assumptions on the estimated outcomes and whether certain specifications improve (or worsen) the statistical fit and the power of the models to explain observed labor supply reactions. In fact, discrete choice labor supply models are often criticized for being a black box due to the numerous explicit and implicit assumptions to be made.

We aim to fill this gap by examining the robustness of structural labor supply models with regard to their empirical setup. In a first step, we, therefore, 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 different empirical specifications which are frequently used in the literature. Second, we set up and estimate 2,219 structural labor supply models each representing a different combination of modeling assumptions. Based on the estimation results we perform a meta-analysis to gather insights how sensitive the statistical fit of the model and the estimated labor supply elasticities are with respect to the underlying assumptions and the empirical specification. In a third step, we introduce a new estimation strategy to estimate preferences and wages jointly. Our approach overcomes the restrictive independence assumptions that has frequently been made in the discrete choice context to facilitate the estimation process.

Our analysis extends existing comparative studies in three ways. First, many of the existing comprehensive surveys in the empirical labor supply literature do not focus on the role of modeling assumptions but either on the principles of alternative estimation strategies (Blundell and MaCurdy, 1999, Blundell et al., 2007) or cross-country comparisons of estimation results (Meghir and Phillips, 2010, Bargain et al., 2012). Second, although many important studies provide sensitivity checks to show that their results are robust with respect to the specification (see, e.g., van Soest, 1995, Euwals and van Soest, 1999, van Soest and Das, 2001, van Soest et al., 2002, Aaberge et al., 2009), they do not take into account all possible combinations of modeling assumption, which is reasonable given the different foci of the papers. Instead, these robustness checks usually narrowed down to a small deviation in just one of the several modeling assumptions. Third, unlike other meta-analyses, such as the one by Evers et al. (2008), who gather estimates from various studies, we first estimate various models on the same data using the same control variables and then assess the results in a meta-regression analysis. Thus, the outcome variables of our meta-analysis, i.e., the statistical fit and the estimated labor supply elasticities, have been derived in a controlled environment.

The results of our meta-analysis confirm previous findings in the literature regarding the insensitivity of the models' performance with respect to the specification of the choice set, the form of the utility function and the inclusion of observed and unobserved preference

heterogeneity, hours restrictions and welfare stigmata. 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 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. 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 new estimation strategy to overcome commonly made but highly restrictive exogeneity assumptions with regard to the wage rate and the labor supply decision.

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 meta-analysis of modeling assumptions and present first results. The new joint 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. As higher working hours increase the consumption but reduce leisure, households face a trade-off between these two goods. Stated mathematically:

$$\max_h U(C, L) = \max_h U(f(wh + I), T - h), \quad (1)$$

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 is equal to disposable income. This setup can also easily be extended to cases with more than one decision maker in the household.

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 usually be the case in presence of the complicated tax and benefits systems in most modern countries. Moreover, restrictive a priori assumptions have to be imposed in order to consistently estimate this kind of model (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 with respect to hours of work. By comparing different levels of utility one avoids the cumbersome maximization process of

Hausman-type models (Aaberge et al., 2009). Moreover, 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. We focus our analysis on the discrete choice approach as it has become a standard procedure in the 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.

$$\begin{aligned} U(C_{ni}, L_i | X_{ni}) &= v(C_{ni}, L_i | X_{ni}) + \epsilon_{ni} \\ U_{ni} &= v_{ni} + \epsilon_{ni} \end{aligned} \quad (2)$$

The utility of household  $n$  from choosing alternative  $i$  is given by  $U_{ni}$ , the observed portion—the *systematic utility function*—is denoted as  $v_{ni}$  and  $\epsilon_{ni}$  denotes an unobserved error term. It is assumed that conditional on random effects incorporated in the model the household's decision satisfies the Independence of Irrelevant Alternatives (IIA) property. In other words, the preference ranking between two choice alternatives does not depend on the presence of a third one.<sup>1</sup> The error terms  $\epsilon$  are i.i.d. and follow the extreme value type I distribution with the cumulative distribution function  $F_\epsilon(v) = \exp(-\exp(-v))$ .

This distributional assumption leads directly to the representation of the labor supply decision as *conditional logit model* (McFadden, 1974):

$$P(U_{ni} > U_{nj}, \forall j \neq i) = \exp(v_{ni}) / \sum_{s=1}^{J_n} \exp(v_{ns}). \quad (3)$$

In order to estimate the preference coefficients, one has to evaluate the systematic utility function  $v$  for every household  $n = 1, \dots, N$  and every choice category within  $j = 1, \dots, J_n$ . 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.

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

---

<sup>1</sup>This assumption may seem rather restrictive at first glance. However, 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.

are possible (see table 1). Stern (1986) discussed the implications and restrictions 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.

There are mainly two ways to include also unobserved heterogeneity, which relaxes the IIA assumption as it allows for additional unobservable variation in preferences between the choice alternatives. In most applied works either a *random coefficient model* (van Soest, 1995) or a *latent class model* (Hoynes, 1996) is assumed. The appropriateness and the performance of both approaches mainly depend on the underlying data.

**Disutility components, fixed costs and hours restrictions** 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 combined with an unobservable stigma that affects the household's utility and prevents the household from taking up benefits (Moffitt, 1983). In the discrete choice context, this can be incorporated by adding dummy variables which capture the potential disutility and doubling the choice alternatives such that the household has to decide between benefit take-up and non-participation in welfare programs (Hoynes, 1996).

Moreover, van Soest (1995) argued that working part-time could be connected with an unobservable disutility as well, because part-time jobs may include 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 rationale remains rather ad hoc. Aaberge et al. (1995) provide a more convenient theoretical framework which delivers a structural interpretation of fixed costs and the disutility connected to certain hours alternatives. In their model, households choose between (latent) job opportunities which may not only differ with regard to the working hours but also in terms of wages, non-monetary attributes and also in terms of their availability. However, when it comes to the empirical specification, the job opportunity framework is largely equivalent to a combination of fixed costs and specific hours restrictions.

**Wage imputation** 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 levels of consumption 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 that has

to be made 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 explicitly ignore or include potential sample selection issues in the observed wages and apply a Heckman correction.

After estimating the wage equation, another important question is whether the potential errors in the wage rate prediction are incorporated 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 prediction will be considerably different from the observed wage. 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. While of course this procedure lacks for a theoretical rational, it reduces the computational burden of the estimation substantially.

### 2.3 Estimation approach

The named extensions especially with regard to the inclusion of unobserved heterogeneity, the incorporation of wage prediction errors and the joint estimation of preferences and wages complicate the estimation procedure and lead to the more general representation as *mixed logit model* (Train, 2009). Thus, taking the most general specification as reference, the likelihood function for the different specifications can be written as:

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

where  $i$  denotes the alternative chosen by individual  $n$ . The likelihood contributions depend not only on the systematic 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 Germany. 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  $\beta_u$ . Similarly, the individual likelihood contributions have to be integrated over the range of possible wage predictions  $\hat{w}_{ni}$ . As both variables will typically not be uniformly distributed, the choice probability has to be weighted by the (joint) probability density function of the random components.

The model as written down in equation (4) is very general, but in turn, it is no longer possible to find an analytical solution to the maximization problem. Instead, Train (2009) proposes

the use of maximum simulated likelihood methods. Thereby, the double integral has to be approximated and averaged over  $r = 1, \dots, R$  random draws from the joint distribution of  $\beta_u$  and  $\hat{w}_{ni}$ . The resulting simulated log-likelihood converges to the log-likelihood of the model as denoted in equation (4) when the number of draws goes to infinity.

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

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).

The representation of the labor supply decision as random utility model allows us to estimate the model without imposing constraints on the coefficients beforehand. Whether or not the estimated coefficients are in line with the economic intuition and thus the assumption of utility maximization can be checked afterwards. Euwals and van Soest (1999) point out that it is only necessary to check that the marginal utility of consumption is non-negative as differences between desired and observed hours of work, e.g., due to constraints on the labor market may justify also negative marginal utilities in leisure.

## 2.4 Existing Literature

Tables 1 and 2 provide an overview on the empirical specification of the 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 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 observed to be concentrated at few hours categories, most authors include either fixed costs of working or hours restrictions or both in their models. As explained before, 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 applied models explicitly allowed for stigma effects and non-take-up of welfare benefits. This is interesting because it is a common finding for modern welfare states that the benefit participation rate deviates substantially from full take-up. Thus, models which do not explicitly account for the potential disutility from welfare participation are expected to clearly over-predict the number of recipients.

Less variation can be found in terms of the model's treatment of wages. While most studies



Table 1: Different model specifications

Paper	Utility	Heterogeneity <sup>a</sup>		Welfare		
	Function	Observed	Unobs.	Choices	Stigma	Dummies
Aaberge et al. (1995, 2009)	Box-Cox	$L$	—	?	—	$FC, HR$
Aaberge et al. (1999, 2000, 2004)	Box-Cox	$L, FC$	—	?	—	$FC, HR$
Dagsvik and Strøm (2006)	Box-Cox	$L, FC$	—	8	—	$FC, HR$
Dagsvik et al. (2011)	Box-Cox	$L, FC$	—	12	—	$FC, HR$
Blundell and Shephard (2012)	Box-Cox	$L, C, S, FC$	$C, S$	6	Yes	$FC$
van Soest (1995)	Translog	$L$	—/ $L^b$	5/6	—	—/ $HR$
Euwals and van Soest (1999)	Translog	$L, FC$	$L$	8/16	—	$FC$
van Soest and Das (2001)	Translog	$L, FC$	$L$	8/15	—	$FC$
Flood et al. (2004)	Translog	$L, L^2, S$	$L, L^2, S$	7	Yes	—
Haan (2006)	Translog	$L, C$	—/ $C$	3, 5 <sup>c</sup>	—	$HR$
Flood et al. (2007)	Translog	$L, C, FC, S$	$L, C, FC, S$	5	Yes	$FC$
Hoynes (1996)	Stone-Geary	$L, S$	$L, S$	3/4	Yes	—/ $FC$
van Soest et al. (2002)	Polynomial	$L$	$L$	16/361	—	$FC$
Keane and Moffitt (1998)	Quadratic	$L, S$	$L, S$	3	Yes	—
Blundell et al. (1999, 2000)	Quadratic	$L, C, FC$	$C, S$	5	Yes	$FC$
Bargain et al. (2012)	Quadratic	$L, C, FC$	$C$	7	—	$FC$

<sup>a</sup>  $L$  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.

<sup>b</sup> Robustness checks and alternative model specifications are separated by slashes.

<sup>c</sup> Three choice alternatives for males, five for females.

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, 2000, 2004)	Simultaneous	—	Full sample	—
Keane and Moffitt (1998)	Simultaneous	—	Non-workers	—
van Soest et al. (2002)	Simultaneous	—	Full sample?	Integrated out?
Blundell and Shephard (2012)	Simultaneous/Two step	—	Non-workers	Integrated out
van Soest (1995)	Two step	Yes	Non-workers	—/Integrated out <sup>b</sup>
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. (2012)	Two step	Yes	Full sample	Single random draw

<sup>b</sup> Robustness checks and alternative model specifications are separated by slashes.

estimate wages and the labor supply decision in a two-step procedure, only the models of Aaberge et al. (1999, 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 process.

### 3 Data

The estimations in this paper are performed on the German Socio-Economic Panel (SOEP), which is a representative household panel survey for Germany (Wagner et al., 2007). We use the 2011 wave of SOEP, which includes data from the year 2010 as well as data on the labor supply behavior and incomes from the preceding year. We focus our analysis on the working age population and thus exclude individuals younger than 17 or above the retirement age of 65. Our sample is further restricted to those households where at least one decision maker 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 (see table 3). First, single men and single women either in a single household or living with dependent children. These households are denoted by SgM and SgF respectively. Our estimation subsample contains 749 households with single males and 1,044 households with single females. In addition, we specify three different kinds of couple households. First, the column denoted by CoM describes 564 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). The column CoF shows that we have 960 couple households where the male partner has a 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. The last column shows that our sample includes 2,926 couple households where both partners are flexible regarding their labor supply behavior. The second row of table 3 denotes the number of choice alternatives we assume the households to face. As said before, we assume the first four household types to face only seven possible labor supply states (either non-participation or working 10, 20, 30, 40, 50 or 60 hours per week),

whereas couples with two flexible partners face  $7^2$  and thus 49 alternatives.

Table 3: Households and choice alternatives

	SgM	SgF	CoM	CoF	CoMF
Households	749	1,044	564	960	2,926
Alternatives	7	7	7	7	49

For the computation of consumption levels for the different choice categories we rely on the policy simulation model of the Institute for the Study of Labor, IZAΨMOD (see Peichl et al., 2010, for a detailed documentation). IZAΨMOD incorporates a very detailed representation of the German tax and benefit system.

For models in which the wage prediction errors are integrated out during the estimation process, we have to apply the tax and benefit system for every possible wage rate for every step of the numerical maximization. This slows down the estimation process substantially. To avoid this cumbersome procedure, we approximate the tax and benefit system for these models by using a second-degree polynomial function that transforms gross earnings into disposable income while controlling for several household characteristics  $x_{ni}$  as well as non-labor incomes  $z_{ni}$ :

$$C_{ni} = \beta_{w,1}wage_{ni} + \beta_{w,2}wage_{ni}^2 + \beta_{w,3}x_{ni}wage_{ni} + \beta_{w,4}x_{ni}wage_{ni}^2 + \beta_x z_{ni}x_{ni} + \eta_{ni}. \quad (6)$$

The resulting  $R^2$  shows a good fit of more than 95 % for all population subgroups.

## 4 Meta-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 explain them mainly by study characteristics. In either case, it is hard to draw general conclusions from the reported results. Detailed and reliable insights about the impact of specific modeling assumptions on the estimated outcomes can only be obtained when the labor supply decision is estimated in a common framework and on the same data set. Otherwise one cannot rule out that the results are driven by the underlying data or differences in the modeling setup between different studies.

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 build the source for a meta-analysis that allows us to determine how sensitive or robust the estimated models are with respect to the specification and the wage imputation procedure.

## 4.1 Analysis setup

To generate a reliable data set for the meta-analysis, we combine several different modeling assumptions and estimate all possible combinations. Table 4 shows the different specifications and the number of estimation results. We estimated 2,219 different model specifications for the five distinct labor supply types which leads us to a total of 11,095 maximum likelihood estimations. However, the sample of our meta-analysis is reduced for two reasons. First, not all models did converge in a reasonable time span as we applied an automatic routine to estimate this large number of models (see table 9). Additionally, some models did converge but clearly rejected the basic assumption of a non-negative marginal utility of consumption for a substantial part of the observations (see table 10). In contrast to the continuous hours approach we did not impose this assumption as a constraint to the maximization problem but checked the economic rational afterwards. Therefore, we decided to drop all estimation results from our meta-analysis which did not converge or which predict more than five percent of the sample to have a negative marginal utility with respect to consumption. We loose between 5 and 24 percent of the estimation results and end up with a sample of 9,344 distinct maximum likelihood estimations. As we perform our analysis separately for the different labor supply groups to allow for heterogeneous effects across these groups, we still have sample sizes between 1,687 and 2,110 distinct model specifications and estimation results for all labor supply types.

Table 4 reads as follows. We estimated 572 distinct models with a Box-Cox transformed utility specification for each of the five labor supply groups. But only 526 estimation results are included in our sample for single males, 449 for single females, 385 for couple households where the male partner has a flexible labor supply but the spouse is inflexible, 460 for couple households with inflexible male partners and 519 for couple households where both partners have a flexible labor supply. Of all estimated models (regardless of the functional form of the utility function), 818 models neglected any kind of hours restrictions or fixed costs, 877 models included part-time restrictions, 524 models accounted for fixed costs of work. 160 models were estimated without any random component, 1,156 models are based on just five Halton sequences, 903 models make use of 50 Halton draws.

**Meta-regressions** In order to make the estimation results comparable, we run simple OLS regressions of the statistical fit and the estimated elasticities on the different modeling assumptions of the models:

$$y_{k,i} = \alpha_k + \beta_{k,u}Utility_i + \beta_{k,s}Stigma_i + \beta_{k,h}Draws_i + \beta_{k,d}HoursRes_i + \beta_{k,o}ObsHet_i + \beta_{k,r}RandHet_i + \beta_{k,w}WageImp_i + \epsilon_{k,i} \quad k \in (ela, aic).$$

With the exception of  $Draws_i$ , which is a continuous variable denoting the number of Halton draws used in the estimation, the other covariates are vectors of dummy variables indicating, e.g., a specific utility function. We measure the statistical fit by the Akaike Information Criteria of the models which is defined as:  $aic_i = 2k_i - 2 \log L_i$ , with  $k_i$  as number of estimated coeffi-

Table 4: Estimated model combinations

Model Parameter	Option	N	By Labor Supply Type <sup>a</sup>				
			SgM	SgF	CoM	CoF	CoMF
Utility function	Box-Cox	572	526	449	385	460	519
	Quadratic	891	869	768	750	665	878
	Translog	756	607	546	552	657	713
Welfare stigma	No	1,265	1,184	978	886	972	1,195
	Yes	954	818	785	801	810	915
Hours restrictions	—	818	676	624	538	601	737
	Fixed costs	524	481	445	414	423	509
	Part-time	877	845	694	735	758	864
Number of Halton draws	—	160	154	132	143	132	160
	5	1,156	1,075	952	921	987	1,138
	50	903	773	679	623	663	812
Observed heterogeneity	—	549	504	505	462	526	532
	in C only	549	491	363	401	254	521
	in L only	549	502	501	469	527	521
	in L, C	572	505	394	355	475	536
Unobserved heterogeneity	—	369	338	306	327	324	361
	in C only	466	442	394	389	373	464
	in L only	466	410	346	319	375	436
	in L, C	465	420	361	356	356	431
	with correl.	453	392	356	296	354	418
Wage imputation	Full sample	1,004	936	842	761	861	958
	Non-workers	1,215	1,066	921	926	921	1,152
Wage prediction error	—	991	946	845	851	865	944
	1 random draw	208	203	158	118	117	201
	Integrated out	1,020	853	760	718	800	965
<b>Total</b>		<b>2,219</b>	<b>2,002</b>	<b>1,763</b>	<b>1,687</b>	<b>1,782</b>	<b>2,110</b>

<sup>a</sup> Estimation results which did not converge or which declared more than five percent of the observations to have a negative marginal utility from consumption are excluded.

coefficients and  $\log L_i$  as log-likelihood of the estimated models. To obtain labor supply elasticities in the discrete choice context, we have to increase the own-wage rates by, e.g., one percent and simulate the labor supply reaction. Thereby we directly obtain the uncompensated own-wage labor supply elasticities. We use aggregated total elasticities as dependent variable, thus, we aggregate the labor supply reactions at the intensive and extensive margin over all affected individuals.

The estimation results of these two meta-regressions can be found in tables 5 (AIC as dependent variable) and 7 (elasticities as dependent variable). Both tables present the marginal impact of certain modeling assumptions controlling for several other specification issues. In addition, we present the partial impact of the specifications in tables 6 and 8.

**Reference Model** The presented coefficients are still difficult to interpret. In both tables, the “constant model” refers a pretty simple setup. We assume a translog utility specification neglecting observed and unobserved heterogeneity in preferences as well as fixed costs of working, hours restrictions or any stigma from welfare participation. Thus, the systematic

utility function for a single-decision-maker household is given by:

$$v_{ni} = \beta_C \log C_{ni} + \beta_{CC} \log(C_{ni})^2 + \beta_L \log L_{ni} + \beta_{LL} \log(L_{ni})^2 + \beta_{CL} \log C_{ni} \log L_{ni}. \quad (7)$$

For actual workers, the observed wage rates are treated as exogenous. For non-workers, the missing wage information is estimated beforehand and predicted without incorporating the wage prediction error in the labor supply estimation. The presented coefficients in tables 5 and 7 thus have to be interpreted in comparison to this simple reference model. One can see at the bottom of table 7 that the reference model yields elasticities of 0.150 for single men and 0.154 for single women. Men in couples have slightly higher elasticities of 0.173, women in couple have with 0.310 the highest elasticities. These findings show a similar picture as the sample mean of estimated elasticities (which is however somewhat larger across all four subgroups) and are well in line with the other findings in the literature.

## 4.2 Estimation Results

In this section, we present the results of our meta-analysis of model specifications. First we show how the different assumptions and model setups increase (or worsen) the statistical fit of the estimated models. While this delivers insights how well the models are able to explain and replicate the observed labor market outcomes, the variable of interest is usually not the statistical fit but the out of sample prediction, e.g., in terms of the implied labor supply elasticities. Therefore, in a second step we present the results of our meta-analysis with regard to the uncompensated labor supply elasticities.

**Goodness of fit** Although the statistical fit is usually not the outcome of highest interest, the results in table 5 show several interesting patterns that are relevant for future applications of labor supply models. First of all, the translog utility specification leads to a slight but significant better fit than the quadratic utility function. While this holds for all labor supply groups, the fit is further improved for single males and single females when using a Box-Cox transformed specification. This confirms the usual finding of a good fit for the Box-Cox transformation. The expected patterns are also found with regard to the implementation of hours restrictions, fixed costs and observed preference heterogeneity. These extensions clearly help to explain the observed behavior. The magnitude of the improvement yields interesting insights as well. Our estimates show that hours restrictions in terms of a lack of part-time jobs is a bigger issue for male than for female workers. This finding is well in line with usual findings for Germany that men typically either work full-time or zero hours. In line with this, we find much more heterogeneity in preferences for women than for men. Surprisingly, the performance of the random coefficients models is pretty bad compared to the computational burden of the estimation. Although they improve the fit by small amounts, the improvement is rather insignificantly. Real improvements are only possible when the model allows for unobserved heterogeneity in preferences for leisure and consumption and the setup allows for correlation between both preference coefficients. Although large in magnitude and statistically

significant, the positive coefficient for models including welfare stigma is barely surprising. As the incorporation of the welfare take-up decision increases the choice set (in fact, it may double the number of choice alternatives), it is obvious that also the log-likelihood becomes more negative and the AIC values increase substantially.

The results with regard to the wage imputation show that these specification decisions are equally important to achieve a good fit of the model. It is rather surprising that integrating the wage prediction error out during the estimation process yields a better fit of the model as it—spoken technically—increases the variation in the underlying data but does not add additional coefficients to explain the variation. In contrast to this, predicting wages not only for non-workers but for the full sample increases the fit significantly. However, this is statistically not surprising as it demonstrates how much of the variation in the data is lost only by using predicted wages for the full sample. Both procedures to incorporate wage prediction errors when predicting for the full sample show that the gain in the log-likelihood of the models vanishes though our results also show that the addition of just one single random draw instead of integrating over a wide range of random draws is a very rough approximation.

More generally, our results show that maybe except for the implementation of hours restrictions there is hardly a single issue in the model setup that guarantees a good fit. Instead, there are several small issues that help to explain the observed labor market outcomes. Compared to the models frequently applied in the literature, the only surprising finding is how rarely the applied models allowed for heterogeneity in preferences for consumption.

**Labor Supply Elasticities** As structural labor supply models are frequently used to estimate labor supply elasticities and labor supply reactions to policy changes, even more important than the statistical fit is whether specific modeling assumptions systematically influence the out-of-sample predictions. By running our meta-analysis also with respect to the estimated elasticities, we provide insights on this question.

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. Especially as most of the estimated effects have different signs across the labor supply groups. However, even though the magnitude of the estimated coefficients is rather small, there are some factors that systematically drive the estimated results and may add up to a significant bias in one direction. E.g., the incorporation of welfare stigma and fixed costs of work makes especially part-time work with supplementary welfare participation less attractive and therefore yields higher elasticities. Thus, our results underline the importance of robustness checks and the calculation of standard errors for the estimated elasticities when estimating labor supply reactions.

Substantially more variation can be found when analyzing the impact of the wage imputation and the handling of the wage prediction errors. Our results hold the important message that this part of the model specification is way more relevant to the estimated outcomes than the specification of the utility function. With regard to the decision whether to use predicted wages for non-workers only or to predict wages for the full sample, our analysis reveals that

the latter procedure roughly doubles the estimated elasticities when the model does not account for the wage prediction error. This substantial bias can be explained by the fact that predicting wages for the full sample reduces the variance of the wage distribution substantially. In order to explain the observed hours distribution, the implied elasticities have to increase. To account for wage prediction errors and to integrate these errors out during the estimation reduces the bias 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.

To summarize, our findings clearly confirm previous findings in the literature. Although the empirical specification of the systematic utility function has an impact on the statistical fit and the estimated elasticities, in particular the latter is rather small. It thus may be justified to rely on simpler model setups when the computational burden is a major concern. Although small in magnitude, the results of our meta-analysis still hold the message that the different modeling assumptions may influence the results and thus, we emphasize the importance of robustness checks when estimating a structural labor supply models.

The majority of applied robustness checks has concentrated on the effects of different utility specifications but usually ignored how the underlying wage treatment may influence the results. We find that a lot more variation in outcomes can be found with regard to these assumptions than with respect to the specification of the utility function. Therefore we conclude that most previous robustness checks have 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 meta-analysis shows that the specification of the systematic utility function has a significantly smaller impact on the estimated labor supply elasticities than the wage imputation procedure and the handling of the wage prediction error. Contrary to this finding, it is common practice in applied labor supply works to estimate preferences and thus the labor supply decision conditional on observed or predicted wages. The wage rates are often estimated beforehand and then treated as exogenous within the labor supply estimation.

$$\log L = \sum_{n=1}^N \sum_{j=1}^{J_n} \log P(U_{nj} > U_{nk}, \forall k \neq j | w_{nj} = \hat{w}_n)^{d_{nj}}, \quad (8)$$

where  $\hat{w}_n$  is predicted beforehand for non-workers and either observed or predicted for actual employees. Obviously, this is a rather strict assumption. Aaberge et al. (1995), Keane and Moffitt (1998) and van Soest et al. (2002) depart from this procedure by estimating wages and



preference and thus the labor supply decision jointly in a simultaneous maximum likelihood setup:

$$\begin{aligned} \log L = & \sum_{n=1}^N \sum_{j=1}^{J_n} \log P(U_{nj} > U_{nk}, \forall k \neq j | w_{nj} = \hat{w}_n)^{d_{nj}} \\ & + \sum_{n \in E} \log \phi \left( \log w_n - \log \widehat{w}_n \right) - \log \sigma_w, \end{aligned} \quad (9)$$

where  $E$  is the subset of employed individuals with wage rate observed,  $\phi(z)$  is the probability density function of a standard normal distribution and  $\sigma_w$  is the standard error of the wage regression, i.e.,  $se(\log w_n - \log \widehat{w}_n)$ . Although the simultaneous estimation is computationally more demanding, it is also clearly more efficient than the two-step procedure. However, even in the simultaneous it is assumed that wages and the labor supply decision, i.e., hours of work, are independent from one another. Only little effort has been taken to relax this strong independence assumption. Blundell and Shephard (2012) further extend the simultaneous maximum likelihood procedure by allowing for correlation between unobserved preference heterogeneity and the wage rates by including both in a mixed logit model and integrating unobserved multivariate normally distributed error terms  $\epsilon_\beta, \epsilon_w$  out:

$$\begin{aligned} \log(SL) = & \sum_{n \in E} \log \left( \frac{1}{N} \sum_{r=1}^R \sum_{j=1}^{J_n} P(U_{nj}^r > U_{nk}^r, \forall k \neq j | w_{nj} = \hat{w}_n, \epsilon_\beta^r, \epsilon_w^r = \log w_n - \log \widehat{w}_n)^{d_{nj}} \right) \\ & + \sum_{n \in U} \log \left( \frac{1}{N} \sum_{r=1}^R \sum_{j=1}^{J_n} P(U_{nj}^r > U_{nk}^r, \forall k \neq j | w_{nj} = \hat{w}_n, \epsilon_\beta^r, \epsilon_w^r)^{d_{nj}} \right) \\ & + \sum_{n \in E} \log \phi \left( \log w_n - \log \widehat{w}_n \right) - \log \sigma_w. \end{aligned} \quad (10)$$

Although this allows for at least some interaction between preferences and wages, it still assumes that the labor supply decision is exogenous to the wage rate. We propose an estimation strategy that overcomes this restrictive assumption and allows wages and hours of work to be correlated. In addition, we use predicted wages for actual workers as well as non-workers:

$$\begin{aligned} \log(SL) = & \sum_{n=1}^N \log \left( \frac{1}{N} \sum_{r=1}^R \sum_{j=1}^{J_n} P(U_{nj}^r > U_{nk}^r, \forall k \neq j | w_{nj} = \hat{w}_{nj}, \epsilon_\beta^r, \epsilon_w^r)^{d_{nj}} \right) \\ & + \sum_{n \in E} \sum_{j=1}^{J_n} \log \phi \left( \log w_{nj} - \log \widehat{w}_{nj} \right)^{d_{nj}} - \log \sigma_w. \end{aligned} \quad (11)$$

*To be continued.*

## 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 that are usually made when applying this kind of discrete choice model impact the statistical fit of the model as well as the estimated outcomes in terms of labor supply reactions.

In this paper, we provide an overview on the most important specification issues and carry out a comprehensive meta-analysis in order to disentangle the driving factors behind modern labor supply models. Our results show that even if the modeling assumptions concerning the specification of the direct utility function increase or worsen the statistical fit, i.e., the power to explain the observed labor supply behavior, the models are rather robust to their concrete specification when it comes to labor supply responses in reaction to changes in wages or policy rules. These results are well in line with previous findings in the literature. In contrast to the robustness regarding the utility function, the models are highly responsive to changes in the underlying wage information and thus with regard to the wage imputation procedure. In fact, our results indicate that, e.g., using predicted wages for the full sample instead of predicting wages for non-workers, roughly doubles the estimated elasticities. 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 estimated outcomes. Surprisingly little effort has been taken so far to check the robustness of the models to these assumptions.

Therefore, we further tackle this issue and provide a new estimation method that overcomes the restrictive independence assumptions previously made in the context of discrete choice models. Our results show... *to be continued.*

## 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**: 403–422.
- Aaberge, R., Colombino, U. and Strøm, S. (2000). Labor supply responses and welfare effects from replacing current tax rules by a flat tax: Empirical evidence from italy, norway and sweden, *Journal of Population Economics* **13**: 595–621.
- Aaberge, R., Colombino, U. and Strøm, S. (2004). Do more equal slices shrink the cake? an empirical investigation of tax-transfer reform proposals in italy, *Journal of Population Economics* **17**: 767–785.
- 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. (2012). Comparing labor supply elasticities in europe and the us: New results, *IZA Discussion Papers* 6735.
- 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, *The Journal of Human Resources* 15(3): 409–423.
- 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: 803–826.
- Euwals, R. and van Soest, A. (1999). Desired and actual labour supply of unmarried men and women in the netherlands, *Labour Economics* 6: 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, *The 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.
- MaCurdy, T. E., Green, D. and Paarsch, H. (1990). Assessing empirical approaches for analyzing taxes and labor supply, *The 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, *The American Economic Review* **73**(5): 1023–1035.
- 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.
- van Soest, A. (1995). Structural models of family labor supply – a discrete choice approach, *The 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.

Table 5: Marginal effects on AIC

depend. var = AIC	(1) SgM	(2) SgF	(3) CoM	(4) CoF	(5) CoMF
Utility function					
Quadratic	27.534 (2.516)	16.700 (3.924)	17.016 (1.160)	2.571 (1.619)	149.697 (10.989)
Box-Cox	-58.009 (3.967)	-71.076 (4.939)	25.122 (2.042)	6.746 (1.609)	29.099 (13.292)
Welfare stigma	210.077 (2.926)	522.271 (3.900)	93.305 (1.315)	68.780 (1.436)	602.839 (10.241)
Number of Halton draws	-0.283 (0.064)	-0.339 (0.085)	-0.076 (0.027)	-0.070 (0.032)	-1.912 (0.223)
Hours restrictions					
Part-time restrictions	-308.538 (3.321)	-266.329 (4.328)	-213.091 (1.464)	-107.605 (1.616)	-2,477.424 (11.442)
Fixed costs	-153.788 (4.037)	-121.090 (4.788)	-121.146 (2.130)	-67.389 (1.695)	-1,628.309 (14.628)
Observed heterogeneity					
in C only	-33.208 (3.865)	-84.891 (5.379)	-18.489 (1.829)	-117.150 (2.436)	-118.578 (14.417)
in L only	-24.458 (3.868)	-115.600 (4.795)	-18.705 (1.708)	-126.942 (1.655)	-316.915 (13.881)
in C and L	-33.986 (3.899)	-136.380 (5.088)	-20.300 (1.747)	-131.309 (1.730)	-345.763 (13.340)
Unobserved heterogeneity					
in C only	-0.450 (4.745)	4.709 (6.100)	7.040 (2.120)	12.732 (1.894)	-11.630 (14.836)
in L only	1.368 (4.781)	5.413 (6.158)	1.189 (1.940)	8.332 (2.024)	12.442 (13.981)
in C and L	-21.743 (4.934)	-13.312 (6.129)	-1.265 (2.108)	-0.071 (2.165)	27.414 (15.677)
in C and L (with correl.)	-34.987 (4.774)	-26.542 (6.114)	-12.415 (2.107)	-3.285 (2.136)	-32.527 (16.182)
Wage imputation					
Full sample, no correction	-86.521 (3.125)	-130.779 (5.082)	-39.806 (1.698)	-99.468 (2.018)	-781.494 (12.739)
Full sample, error integrated out	-51.619 (3.668)	-78.712 (5.115)	-10.761 (1.914)	-17.288 (2.075)	-155.735 (16.828)
Full sample, 1 random draw	-6.448 (3.790)	18.196 (5.640)	1.322 (3.047)	27.086 (2.349)	84.357 (18.345)
Non-workers, error integrated out	-40.973 (4.480)	6.840 (5.238)	11.879 (1.767)	0.258 (1.654)	-249.230 (12.182)
Constant	2,631.527 (5.116)	3,774.335 (6.724)	1,870.171 (2.542)	3,552.247 (2.744)	21,122.255 (18.565)
N	2,002	1,763	1,687	1,782	2,110
R <sup>2</sup>	0.892	0.943	0.945	0.919	0.965
Mean	2,469.915	3,727.804	1,781.995	3,413.864	19,573.604
SD	186.851	313.322	110.033	97.537	1,185.080
Minimum	2,028.757	3,196.046	1,597.286	3,187.719	17,366.828
Maximum	3,024.334	4,519.104	2,033.445	3,676.718	22,383.174
P1	2,057.551	3,241.967	1,612.034	3,205.656	17,555.791
P99	2,955.236	4,412.003	2,011.062	3,649.208	21,913.861

Notes: Robust standard errors, clustered by models, in parentheses.

Table 6: Partial effects on AIC

depend. var = AIC	(1) SgM	(2) SgF	(3) CoM	(4) CoF	(5) CoMF
Utility function					
Translog	-11.837 (8.363)	24.269 (16.040)	-31.366 (5.286)	-0.976 (4.597)	-108.903 (53.367)
Quadratic	86.716 (8.380)	91.918 (15.003)	30.253 (5.472)	1.573 (5.073)	260.047 (53.569)
Box-Cox	-97.060 (8.897)	-146.388 (15.735)	-3.209 (6.024)	-0.735 (4.957)	-209.292 (54.401)
Welfare stigma	205.747 (7.303)	536.612 (8.092)	95.416 (4.836)	64.200 (4.402)	505.004 (50.859)
Number of Halton draws	-0.381 (0.187)	-0.900 (0.333)	-0.254 (0.122)	-0.213 (0.105)	-5.241 (1.160)
Hours restrictions					
None	236.071 (7.161)	184.386 (15.171)	178.031 (3.590)	89.645 (4.348)	2,127.538 (26.280)
Part-time restrictions	-250.373 (6.072)	-239.678 (13.799)	-170.634 (3.259)	-86.141 (4.166)	-1,859.884 (29.907)
Fixed costs	45.342 (7.743)	79.732 (16.437)	17.729 (4.472)	5.625 (5.007)	-184.820 (39.366)
Observed heterogeneity					
None	36.630 (9.611)	101.908 (16.242)	24.143 (5.965)	135.868 (3.836)	290.181 (58.934)
in C only	-11.346 (9.748)	49.068 (18.864)	-5.487 (6.207)	-60.089 (5.440)	123.995 (59.753)
in L only	0.486 (9.598)	-62.585 (16.136)	-8.401 (5.965)	-44.727 (4.376)	-172.652 (59.102)
in C and L	-25.931 (9.504)	-92.897 (17.418)	-12.763 (6.648)	-59.371 (4.451)	-241.001 (58.848)
Unobserved heterogeneity					
None	8.412 (11.200)	13.171 (20.319)	3.940 (6.700)	0.871 (5.737)	26.822 (67.110)
in C only	19.050 (10.146)	17.606 (17.882)	17.036 (6.352)	13.182 (5.746)	41.890 (62.290)
in L only	11.731 (10.393)	33.917 (18.905)	-4.286 (7.022)	3.580 (5.697)	1.372 (63.510)
in C and L	-9.881 (10.315)	-16.537 (18.247)	-1.509 (6.608)	-6.026 (5.814)	26.277 (65.027)
in C and L (with correl.)	-30.039 (10.150)	-47.172 (18.104)	-18.869 (6.764)	-12.205 (5.810)	-97.483 (64.846)
Wage imputation					
Full sample imputation	-70.193 (8.098)	-223.245 (13.827)	-50.495 (5.242)	-63.364 (4.399)	-298.669 (51.692)
Error integrated out	-16.467 (8.670)	0.614 (15.612)	16.265 (5.417)	31.055 (4.526)	63.975 (51.650)
Full sample, no correction	-51.777 (9.326)	-127.543 (15.127)	-44.673 (6.361)	-103.475 (5.029)	-636.284 (66.563)
Full sample, error integrated out	-33.090 (10.217)	-125.272 (19.925)	-20.837 (7.147)	-0.257 (5.548)	76.119 (67.627)
Full sample, 1 random draw	-51.250 (11.017)	-200.023 (13.741)	-32.422 (9.647)	33.737 (8.141)	126.429 (84.811)
Non-workers, error integrated out	3.721 (11.003)	90.289 (18.830)	36.175 (6.021)	40.856 (4.835)	24.328 (56.507)
N	2,002	1,763	1,687	1,782	2,110
Mean	2,469.915	3,727.804	1,781.995	3,413.864	19,573.604
SD	186.851	313.322	110.033	97.537	1,185.080
Minimum	2,028.757	3,196.046	1,597.286	3,187.719	17,366.828
Maximum	3,024.334	4,519.104	2,033.445	3,676.718	22,383.174
P1	2,057.551	3,241.967	1,612.034	3,205.656	17,555.791
P99	2,955.236	4,412.003	2,011.062	3,649.208	21,913.861

Notes: Robust standard errors, clustered by models, in parentheses.

Table 7: Marginal effects on elasticities

depend. var = elasticities	(1) SgM	(2) SgF	(3) CoM	(4) CoF
Utility function				
Quadratic	0.001 (0.002 )	-0.021 (0.003 )	0.016 (0.002 )	0.020 (0.003 )
Box-Cox	0.002 (0.002 )	0.001 (0.003 )	0.026 (0.002 )	0.046 (0.003 )
Welfare stigma	0.004 (0.002 )	0.048 (0.003 )	0.015 (0.002 )	0.018 (0.003 )
Number of Halton draws	-0.000 (0.000 )	-0.000 (0.000 )	-0.000 (0.000 )	-0.000 (0.000 )
Hours restrictions				
Part-time restrictions	-0.007 (0.002 )	0.009 (0.003 )	0.032 (0.002 )	0.026 (0.003 )
Fixed costs	0.013 (0.002 )	0.021 (0.003 )	0.044 (0.002 )	0.039 (0.003 )
Observed heterogeneity				
in C only	-0.014 (0.002 )	-0.017 (0.003 )	0.002 (0.002 )	0.006 (0.004 )
in L only	-0.010 (0.002 )	-0.016 (0.003 )	0.001 (0.002 )	-0.026 (0.003 )
in C and L	-0.018 (0.002 )	-0.032 (0.004 )	-0.002 (0.002 )	-0.026 (0.003 )
Unobserved heterogeneity				
in C only	-0.002 (0.002 )	-0.006 (0.004 )	0.001 (0.002 )	0.000 (0.004 )
in L only	0.001 (0.002 )	-0.007 (0.004 )	-0.020 (0.002 )	-0.011 (0.004 )
in C and L	0.013 (0.003 )	0.016 (0.004 )	-0.030 (0.003 )	-0.023 (0.004 )
in C and L (with correl.)	0.009 (0.003 )	0.006 (0.004 )	-0.029 (0.003 )	-0.024 (0.003 )
Wage imputation				
Full sample, no correction	0.150 (0.002 )	0.200 (0.004 )	0.215 (0.002 )	0.344 (0.003 )
Full sample, error integrated out	0.086 (0.003 )	0.148 (0.005 )	0.060 (0.003 )	0.075 (0.004 )
Full sample, 1 random draw	0.037 (0.002 )	0.040 (0.003 )	0.023 (0.003 )	-0.049 (0.004 )
Non-workers, error integrated out	-0.036 (0.002 )	-0.034 (0.003 )	-0.010 (0.002 )	-0.014 (0.003 )
Constant	0.150 (0.003 )	0.154 (0.005 )	0.173 (0.003 )	0.310 (0.005 )
N	2,002	1,763	1,660	1,718
R <sup>2</sup>	0.797	0.781	0.884	0.900
Mean	0.176	0.218	0.251	0.418
SD	0.078	0.107	0.094	0.153
Minimum	-0.082	-0.053	0.024	-0.198
Maximum	0.382	0.521	0.480	0.789
P1	0.019	0.052	0.104	0.223
P99	0.344	0.501	0.465	0.761

Notes: Robust standard errors, clustered by models, in parentheses.

Table 8: Partial effects on elasticities

depend. var = elasticities	(1) SgM	(2) SgF	(3) CoM	(4) CoF
Utility function				
Translog	0.002 (0.004 )	0.009 (0.006 )	-0.004 (0.005 )	-0.026 (0.007 )
Quadratic	0.012 (0.003 )	0.008 (0.005 )	0.005 (0.005 )	0.033 (0.008 )
Box-Cox	-0.018 (0.004 )	-0.021 (0.005 )	-0.003 (0.005 )	-0.009 (0.008 )
Welfare stigma	-0.004 (0.004 )	0.024 (0.005 )	0.006 (0.005 )	0.010 (0.007 )
Number of Halton draws	-0.001 (0.000 )	-0.001 (0.000 )	-0.001 (0.000 )	-0.001 (0.000 )
Hours restrictions				
None	0.003 (0.004 )	-0.010 (0.005 )	-0.028 (0.005 )	-0.016 (0.008 )
Part-time restrictions	-0.008 (0.004 )	0.009 (0.005 )	0.017 (0.005 )	0.006 (0.007 )
Fixed costs	0.008 (0.004 )	-0.000 (0.006 )	0.010 (0.005 )	0.011 (0.008 )
Observed heterogeneity				
None	0.012 (0.004 )	0.009 (0.006 )	-0.004 (0.005 )	-0.005 (0.008 )
in C only	-0.003 (0.004 )	0.016 (0.006 )	0.008 (0.005 )	0.097 (0.011 )
in L only	-0.001 (0.004 )	-0.014 (0.006 )	-0.002 (0.005 )	-0.036 (0.008 )
in C and L	-0.009 (0.004 )	-0.010 (0.006 )	-0.001 (0.006 )	-0.017 (0.008 )
Unobserved heterogeneity				
None	-0.006 (0.004 )	-0.007 (0.006 )	0.011 (0.005 )	-0.002 (0.009 )
in C only	-0.014 (0.004 )	-0.014 (0.006 )	0.016 (0.005 )	0.013 (0.009 )
in L only	-0.006 (0.004 )	-0.009 (0.006 )	-0.004 (0.006 )	-0.002 (0.009 )
in C and L	0.014 (0.005 )	0.023 (0.007 )	-0.015 (0.006 )	-0.003 (0.010 )
in C and L (with correl.)	0.012 (0.005 )	0.007 (0.007 )	-0.011 (0.006 )	-0.007 (0.010 )
Wage imputation				
Full sample imputation	0.119 (0.002 )	0.151 (0.004 )	0.128 (0.004 )	0.184 (0.006 )
Error integrated out	-0.045 (0.003 )	-0.050 (0.005 )	-0.060 (0.004 )	-0.108 (0.006 )
Full sample, no correction	0.139 (0.002 )	0.172 (0.005 )	0.198 (0.003 )	0.329 (0.003 )
Full sample, error integrated out	0.056 (0.004 )	0.088 (0.005 )	0.006 (0.004 )	-0.016 (0.006 )
Full sample, 1 random draw	-0.002 (0.003 )	-0.050 (0.004 )	-0.039 (0.004 )	-0.152 (0.006 )
Non-workers, error integrated out	-0.101 (0.002 )	-0.126 (0.003 )	-0.081 (0.003 )	-0.127 (0.005 )
N	2,002	1,763	1,660	1,718
Mean	0.176	0.218	0.251	0.418
SD	0.078	0.107	0.094	0.153
Minimum	-0.082	-0.053	0.024	-0.198
Maximum	0.382	0.521	0.480	0.789
P1	0.019	0.052	0.104	0.223
P99	0.344	0.501	0.465	0.761

Notes: Robust standard errors, clustered by models, in parentheses.



Table 9: Estimation results—Convergence probabilities of the different models

	(1)	(2)	(3)	(4)	(5)
	SgM	SgF	CoM	CoF	CoMF
converged					
Utility function					
Quadratic	1.223 <sup>*</sup> (0.496)	-0.176 (0.388)	2.399 <sup>*</sup> (1.112)	0.186 (0.350)	2.224 <sup>***</sup> (0.547)
Box-Cox	-2.708 <sup>***</sup> (0.514)	-2.620 <sup>***</sup> (0.392)	-5.501 <sup>***</sup> (0.588)	-1.354 <sup>***</sup> (0.373)	-3.664 <sup>***</sup> (0.527)
Number of Halton draws	-0.107 <sup>***</sup> (0.0160)	-0.0826 <sup>***</sup> (0.0106)	-0.0574 <sup>***</sup> (0.00688)	-0.0892 <sup>***</sup> (0.0139)	-0.115 <sup>***</sup> (0.0144)
Hours restrictions					
Part-time restrictions	2.885 <sup>***</sup> (0.613)	1.402 <sup>***</sup> (0.362)	2.176 <sup>***</sup> (0.359)	2.000 <sup>***</sup> (0.418)	4.318 <sup>***</sup> (0.748)
Fixed costs	1.066 <sup>*</sup> (0.418)	0.921 <sup>*</sup> (0.375)	-0.361 (0.275)	0.600 (0.329)	0.936 <sup>**</sup> (0.333)
het_obson_c	-1.247 <sup>*</sup> (0.541)	-0.367 (0.405)	0.897 <sup>*</sup> (0.425)	0.163 (0.400)	-0.433 (0.372)
het_obson_l	-1.006 (0.596)	-0.367 (0.433)	-0.818 <sup>*</sup> (0.373)	0.251 (0.410)	-1.268 <sup>***</sup> (0.368)
het_obson_cl	0.140 (0.714)	-0.425 (0.560)	-1.818 <sup>***</sup> (0.515)	-0.938 (0.542)	0.502 (0.555)
het_randon_c	0.0687 (0.394)	0.347 (0.314)	0.632 (0.667)	-11.80 <sup>***</sup> (1.013)	-13.22 <sup>***</sup> (0.962)
o.het_randon_l	0 (.)	0 (.)	-2.743 <sup>***</sup> (0.492)	-15.26 <sup>***</sup> (0.329)	-16.94 <sup>***</sup> (0.677)
o.het_randon_cl	0 (.)	0 (.)	0.600 (0.743)	11.69 <sup>***</sup> (1.062)	12.83 <sup>***</sup> (1.025)
het_randon_corr	-0.162 (0.416)	-0.269 (0.340)	-0.855 <sup>*</sup> (0.360)	0.253 (0.337)	0.113 (0.361)
Full sample imputation	0.259 (0.465)	-0.508 (0.336)	-1.309 <sup>***</sup> (0.307)	0.245 (0.349)	-1.217 <sup>***</sup> (0.344)
Single random draw	-0.605 (0.667)	-0.0633 (0.523)	-0.0810 (0.428)	-0.558 (0.507)	0.968 (0.732)
Error integrated out	-1.089 <sup>***</sup> (0.324)	-1.336 <sup>***</sup> (0.267)	-0.447 (0.269)	-0.816 <sup>**</sup> (0.278)	-0.398 (0.267)
Constant	8.905 <sup>***</sup> (1.182)	7.605 <sup>***</sup> (0.741)	10.24 <sup>***</sup> (0.956)	21.78 <sup>***</sup> (0.893)	25.65 <sup>***</sup> (1.295)
Observations	1384	1384	2219	2219	2219
R <sup>2</sup>					

Standard errors in parentheses

Notes: Robust standard errors, clustered by models, in parentheses.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 10: Marginal effects on dudes

depend. var = dudes	(1)	(2)	(3)	(4)	(5)
	SgM	SgF	CoM	CoF	CoMF
Utility function					
Quadratic	-0.042 (0.005)	-0.021 (0.003)	-0.026 (0.005)	0.018 (0.002)	-0.000 (0.001)
Box-Cox	-0.041 (0.004)	-0.016 (0.002)	-0.027 (0.004)	0.002 (0.001)	-0.006 (0.001)
Welfare stigma	0.036 (0.004)	0.008 (0.002)	0.000 (0.004)	-0.002 (0.001)	-0.001 (0.001)
Number of Halton draws	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)
Hours restrictions					
Part-time restrictions	-0.030 (0.004)	-0.007 (0.003)	-0.035 (0.005)	-0.013 (0.002)	-0.005 (0.001)
Fixed costs	-0.025 (0.005)	-0.010 (0.003)	-0.034 (0.005)	-0.007 (0.002)	-0.003 (0.001)
Observed heterogeneity					
in C only	-0.000 (0.005)	0.029 (0.002)	0.007 (0.005)	0.060 (0.002)	0.002 (0.001)
in L only	0.001 (0.006)	0.002 (0.003)	-0.010 (0.006)	0.000 (0.001)	-0.001 (0.001)
in C and L	0.001 (0.005)	0.029 (0.003)	0.015 (0.006)	0.016 (0.002)	0.000 (0.001)
Unobserved heterogeneity					
in C only	-0.007 (0.005)	-0.003 (0.002)	0.005 (0.003)	0.006 (0.002)	-0.002 (0.001)
in L only	0.002 (0.006)	0.006 (0.003)	0.022 (0.005)	0.003 (0.002)	0.001 (0.001)
in C and L	-0.001 (0.005)	0.000 (0.002)	0.012 (0.004)	0.010 (0.002)	0.000 (0.001)
in C and L (with correl.)	0.015 (0.007)	0.006 (0.004)	0.040 (0.006)	0.010 (0.002)	0.004 (0.002)
Wage imputation					
Full sample, no correction	0.003 (0.002)	-0.011 (0.002)	-0.001 (0.002)	-0.021 (0.002)	-0.001 (0.002)
Full sample, error integrated out	0.030 (0.004)	0.017 (0.003)	0.026 (0.004)	-0.004 (0.001)	0.011 (0.001)
Full sample, 1 random draw	0.019 (0.003)	0.006 (0.003)	0.024 (0.004)	0.027 (0.003)	0.001 (0.001)
Non-workers, error integrated out	0.046 (0.006)	0.010 (0.003)	0.046 (0.006)	0.016 (0.002)	0.004 (0.001)
Constant	0.028 (0.008)	0.019 (0.004)	0.023 (0.006)	-0.008 (0.003)	0.004 (0.002)
N	2,159	2,141	2,085	2,149	2,142
R <sup>2</sup>	0.162	0.172	0.165	0.504	0.130
Mean	0.026	0.029	0.036	0.024	0.006
SD	0.092	0.050	0.090	0.044	0.018
Minimum	0.000	0.000	0.000	0.000	0.000
Maximum	0.904	0.721	1.000	0.341	0.595
P1	0.000	0.000	0.000	0.000	0.000
P99	0.542	0.201	0.362	0.198	0.057

Notes: Robust standard errors, clustered by models, in parentheses.