

DISCUSSION PAPER SERIES

IZA DP No. 11425

**The Sensitivity of Structural Labor Supply
Estimations to Modeling Assumptions**

Max Löffler
Andreas Peichl
Sebastian Siegloch

MARCH 2018

DISCUSSION PAPER SERIES

IZA DP No. 11425

The Sensitivity of Structural Labor Supply Estimations to Modeling Assumptions

Max Löffler

University of Cologne and IZA

Andreas Peichl

ifo Institute, University of Munich, CESifo, IZA and ZEW

Sebastian Sieglösch

University of Mannheim, IZA, ZEW and CESifo

MARCH 2018

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

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

ABSTRACT

The Sensitivity of Structural Labor Supply Estimations to Modeling Assumptions*

There is still considerable dispute about the magnitude of labor supply elasticities. While differences in estimates especially between micro and macro models are recently attributed to frictions and adjustment costs, we show that the variation in elasticities derived from structural labor supply models can also be explained by modeling assumptions. Specifically, we estimate 3,456 different models on the same data each representing a plausible combination of frequently made choices. While many modeling assumptions do not systematically affect labor supply elasticities, our controlled meta-analysis shows that results are very sensitive to the treatment of hourly wages in the estimation. For example, different (sensible) choices concerning the modeling of the underlying wage distribution and especially the imputation of (missing) wages lead to point estimates of elasticities between 0.2 and 0.65. We hence conclude that researchers should pay more attention to the robustness of their estimations with respect to the wage treatment.

JEL Classification: C25, C52, H31, J22

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

Corresponding author:

Andreas Peichl
ifo Institute
Ludwig-Maximilians-University Munich
Poschingerstr. 5
81679 Munich
Germany
E-mail: peichl@ifo.de

* The authors thank Rolf Aaberge, Felix Bierbrauer, Richard Blundell, John Dagsvik, Alan Duncan, Peter Haan, Jan Kabátek, Nico Pestel, and Eric Sommer, as well as seminar participants at CPB, IIPF, IMA, IZA, Young Scholar GSOEP Symposium, Statistics Norway, and Verein für Socialpolitik for helpful comments and suggestions on earlier versions.

1 Introduction

Knowing the size of labor supply responses to wage or policy changes has important implications for welfare analysis (Eissa et al., 2008) and optimal taxation (Diamond and Saez, 2011, Immervoll et al., 2011). Despite a long and comprehensive empirical literature on labor supply behavior, there is still substantial variation in the estimated elasticities (see, e.g., Heckman, 1993, Evers et al., 2008, Chetty et al., 2011, Keane and Rogerson, 2012). Potential reasons include differences in preferences, norms, and institutions across countries and over time. But even for the same country, the same period, and the same estimation approach there is still considerable heterogeneity in individuals' estimated responsiveness to wages (Bargain and Peichl, 2016). One explanation for these remaining differences is the use of different and/or wrongly specified empirical models.

In this paper, we aim to investigate this channel by thoroughly scrutinizing state-of-the-art micro labor supply models and their functioning.¹ Structural models are repeatedly criticized for the large number of assumptions and the even larger number of possible combinations of these assumptions (Keane, 2010, Manski, 2014). We test whether the numerous modeling choices actually affect estimated elasticities. More specifically, we check the internal validity of such models by running controlled experiments: we set up and estimate 3,456 different models, each representing a different (plausible) combination of commonly made assumptions. We use two different micro data sets – one for Germany and one for the US – and estimate these different models for five distinct population groups, leading to 17,280 maximum likelihood estimations for each data set. Based on the estimation results, we gather insights into how robust the statistical fit of the models and the estimated labor supply elasticities are with respect to the underlying assumptions.

The modeling assumptions can be categorized in three broad areas. First, researchers need to specify the utility function. This concerns the functional form, its flexibility with respect to observed and unobserved preference heterogeneity, and the inclusion of stigma costs from welfare participation. Second, there are different ways to construct the choice set and to model the availability of job or hours alternatives and fixed costs of working. The third area relates to the treatment of the underlying wage distribution, namely, the imputation of wages for non-workers, whether to use predicted wages for non-workers only or to impute wages for the full sample, and the handling of the wage prediction error when imputing wages. While the second issue has been surveyed in Aaberge et al. (2009), the literature is rather silent on the first and the last area, which is where this paper intends to break new ground. In particular, the treatment of wages has received hardly any discussion in many existing studies.

Our results show that the models' predictions are strongly driven by the treatment of wages in the estimation. For instance, the choice between predicting wage rates for non-workers with missing wage information only or for the full sample – both procedures are

¹ We focus on structural labor supply models which can be used for policy simulations. In addition, several reduced-form approaches are used in the literature to estimate labor supply responses (see Chetty et al., 2011, and Bargain and Peichl, 2016, for recent surveys).

often used in the literature – may double the estimated labor supply elasticities, raising the average own-wage elasticity in our meta-analysis from 0.23 to 0.46. While the former option presumes that individuals optimize with respect to their current wage, the latter specification assumes that all individuals base their labor supply decision on expected wages as derived from the Mincerian wage equation. The handling of wage prediction errors is equally important. Using predicted wages for all individuals but ignoring the forecast error yields an average elasticity of 0.65 as opposed to 0.35 when accounting for the prediction error. In contrast, it turns out that other modeling choices hardly affect the estimated results. Elasticities are largely robust to the specification of the functional form of the utility function, the inclusion of observed or unobserved preference heterogeneity, as well as the modeling of hours restrictions or stigma costs of welfare participation.

We conclude that the attention of previous sensitivity analyses has been mainly concentrated on less important factors while the main driving forces have been neglected, i.e., the interactions between wages, working hours, and preferences. This finding is even more relevant given that most existing models (implicitly) assume exogeneity between the wage equation and the labor supply decision.² Our findings have important policy implications as labor supply elasticities are key parameters when evaluating or designing optimal tax benefit policies. For instance, Diamond and Saez (2011) use an elasticity of 0.25 to derive an optimal top marginal tax rate of 72.7 percent. However, an elasticity of 0.65, as often found when using alternative wage imputation procedures, reduces the optimal tax rate to 50.6 percent, bringing it closer to actually observed values.

Our analysis makes two important contributions to the literature on labor supply estimation. First, there is little evidence on the functioning of structural labor supply models in general. Moreover, if such studies exist, different models are not estimated on the same data set. Existing surveys and meta-analyses focus on either the principles of alternative estimation strategies (Blundell and MaCurdy, 1999, Evers et al., 2008) or cross-country comparisons of empirical findings (Bargain et al., 2014). Robustness checks in previous studies usually limit themselves to small deviations in one or only few of the numerous modeling assumptions. In that respect, we run a controlled meta-analysis, isolating the impact of the model assumptions on estimation outcomes. Second, our analysis points to a hitherto neglected factor that strongly influences the estimated labor elasticities: we show that the treatment of wages in labor supply estimations, which is rarely theoretically motivated nor subject to robustness checks, crucially affects the estimation results.

The remainder of this paper is organized as follows. Section 2 presents the modeling framework and a short overview of the existing literature. Section 3 provides information on the used data and the modeling of the tax and benefit system. In Section 4 we conduct our analysis of modeling assumptions and present the results. Section 5 concludes.

² Only little effort has been made thus far in the context of discrete choice labor supply models to overcome this assumption. Aaberge et al. (1995), Breunig et al. (2008), and Blundell and Shephard (2012) estimate preferences and wages simultaneously, in part also allowing for correlation.

2 Model and Existing Literature

The use of structural discrete choice labor supply estimations has become a standard procedure in the empirical analysis of labor supply for both econometricians and policy makers (see, e.g., the overview in Bargain and Peichl, 2016). The first generation of labor supply models relied on the assumption that the household's utility is maximized over a continuous set of working hours – known as *Hausman approach* (see Hausman, 1981). This approach has been criticized for three reasons: (i) because the consistent estimation relies on rather restrictive *a priori* assumptions (see, e.g., MaCurdy et al., 1990, or Bloemen and Kapteyn, 2008, for details); (ii) the procedure has proven cumbersome when the budget set is non-convex, which will often be the case in presence of complicated tax and benefits systems in most countries; (iii) it has been shown that the estimated elasticities are very sensitive to the underlying wages (Ericson and Flood, 1997, Eklöf and Sacklén, 2000).

Partly motivated by these shortcomings, it has become increasingly popular to model the labor supply decision as the choice between a (finite) set of utility levels instead of deriving the marginal utility. Starting with the works by Aaberge, Dagsvik, and Strøm (1995), van Soest (1995), and Hoynes (1996), a wide range of different empirical specifications of these *discrete choice models* has been applied. For many institutional settings, 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., 2014). Comparing different levels of utility avoids also the cumbersome maximization process of Hausman-type models. We focus our analysis on the discrete choice approach, given that it has become the standard procedure in the literature.

2.1 General Model

Structural labor supply estimations build on the assumption of the well-known neo-classical labor supply model that decision makers maximize their utility by choosing the optimal amount of working hours (or, more generally, the optimal job) subject to a budget constraint. Utility is defined as a function of consumption C_{nj} , leisure L_j , and idiosyncratic preferences for certain jobs, which we denote by ϵ_{nj} . Individual n faces the decision between a set of jobs $j \in J_n$ with working hours h_j and wages w_{nj} , including non-participation,³ and maximizes her utility over job alternatives:

$$\max_{j \in J_n} U(C_{nj}, L_j, \epsilon_{nj}) = \max_{j \in J_n} U(f[w_{nj}h_j, I_n | \mathbf{x}_{nj}], T - h_j, \epsilon_{nj}) \quad (1)$$

where leisure L_j is denoted as difference between the total time endowment T and working hours h_j . Consumption C_{nj} depends on working hours, the hourly wage rate w_{nj} , non-labor income I_n , household and job characteristics \mathbf{x}_{nj} , and the tax benefit system $f[\cdot]$. We assume a static model, which implies that consumption equals disposable income.

³ We denote non-participation as job alternative $j = 0$ with $h_0 = 0$ and $w_{n0} = 0$.

Individuals' true utility is only partly observable to the researcher while idiosyncratic components captured in ϵ_{nj} are latent. We rewrite the utility of individual n choosing job type j accordingly as:

$$U(C_{nj}, L_j, \epsilon_{nj} | \mathbf{x}_{nj}, \boldsymbol{\beta}_n, \boldsymbol{\gamma}_j) = \varphi(C_{nj}, L_j | \mathbf{x}_{nj}, \boldsymbol{\beta}_n, \boldsymbol{\gamma}_j) + \epsilon_{nj} \quad (2)$$

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

$$P(U_{ni} > U_{nj}, \forall j \neq i | \mathbf{x}_n, \boldsymbol{\beta}_n, \boldsymbol{\gamma}) = \frac{\exp(\varphi[C_{ni}, L_i | \mathbf{x}_{ni}, \boldsymbol{\beta}_n, \boldsymbol{\gamma}_i])}{\sum_{s \in J_n} \exp(\varphi[C_{ns}, L_s | \mathbf{x}_{ns}, \boldsymbol{\beta}_n, \boldsymbol{\gamma}_s])} \quad (3)$$

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

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

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

2.2 Identification

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

It is crucial to impose a specific functional form for both $v(C_{nj}, L_{nj})$ and $g(j)$ to obtain consistent estimates of $\boldsymbol{\beta}_n$ and $\boldsymbol{\gamma}_j$. van Soest et al. (2002) show that semi-parametric specifications also yield consistent results. As consumption is a function of working hours and thus leisure, identification of preference parameters relies on (i) the variation

in working hours h_j , hourly wages w_{nj} , non-labor income I_n , and other characteristics \mathbf{x}_{nj} , and (ii) the fact that the tax function $f(w_{nj}h_j, I_n)$ is highly non-linear in h_j and w_{nj} . This also implies that labor market conditions γ_j can only be separated and identified on the assumption of a specific functional form (Dagsvik and Strøm, 2006).

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

$$E[\beta_n w_{nj} | \mathbf{x}_{nj}] = 0 \qquad E[\gamma_j w_{nj} | \mathbf{x}_{nj}] = 0 \qquad (5)$$

The main reason for this assumption is that it reduces the computational burden substantially and makes the estimation more convenient.

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

2.3 Modeling Decisions

Choice Set. The first modeling decision relates to the construction of the choice set. Most authors simply pick a set of representative levels of hours of work and assume (small) identical choice sets for the whole population. In our analysis, we follow this literature and assume that households with a single decision maker face seven possible labor supply states, i.e., either non-participation or working 10, 20, 30, 40, 50 or 60 hours per week. Couple households are assumed to face 7^2 alternatives. The results are generally not sensitive to the number of choices (e.g., 4 vs. 7 vs. 13) or the exact value assigned to each category (see, e.g., Bargain et al., 2014). As noted before, we focus on other aspects of the model set up, namely the specification of the utility function and the treatment of wages. See Aaberge et al. (2009) for a detailed discussion of alternative representations of the choice set.

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 $v(\cdot)$ represents the direct utility function of the household. Most applications rely on either a translog, a quadratic or a Box-Cox transformed utility specification. However, several other choices are possible.

Heterogeneity in Preferences. Heterogeneity in the labor supply behavior along observable characteristics can be rather easily introduced in the context of structural labor supply models by extending the utility specification. The preference coefficients of the di-

rect utility function are usually interacted with some observed household characteristics, such as age or the presence of children, as taste shifters.

Additionally accounting for unobserved heterogeneity overcomes the IIA assumption as it allows for unobservable variation in preferences between choice alternatives. There are two main ways to do so: in most applied works, either a *random coefficient model* (van Soest, 1995) or a *latent class model* (Hoynes, 1996) is assumed. The former typically assumes a set of preference coefficients to be (multivariate) normally distributed, whereas the latter allows 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 the literature.

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

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

2.4 Wage Imputation Procedure

In addition to the specification of the utility function, there are important modeling assumptions regarding the wage imputation. In order to calculate the disposable income for the different choice alternatives, one needs information on the hourly wage rates. While for actual workers the wage rate can be calculated by gross earnings and hours of work (we use standardized working hours to reduce the potential division bias, see Borjas, 1980, and Ziliak and Kniesner, 1999, for a discussion), the wage information is typically missing for non-workers. The first decision is how to deal with missing wages in the estimation process. In practice, wages are either estimated beforehand and treated

as given within the estimation of the labor supply model or wages and preferences are estimated jointly. In addition, one has to decide whether the estimated wage rates are used only if wages are not observed or for the full sample (see MaCurdy et al., 1990, for a discussion of the advantages and disadvantages of both approaches). In either case, one can ignore or explicitly account for potential sample selection issues in observed wages.

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

2.5 Estimation Approach

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

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

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

The model as shown in equation (6) is very general and less restrictive than the conditional logit set-up. In turn, it is no longer possible to find an analytical solution. Train (2009) proposes the use of maximum simulated likelihood methods instead. In order to retrieve the simulated likelihood, the double integral has to be approximated

and averaged over $r = 1, \dots, R$ random draws from the distributions of $(\boldsymbol{\beta}_n, \gamma)$ and \hat{w}_{nj} . The simulated log-likelihood is given by:

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

When the number of draws goes to infinity, the simulated log-likelihood in (7) converges to the log-likelihood of the model denoted in (6). Instead of relying on conventional random draws, we approximate the likelihood function using pseudo-random Halton sequences. This reduces the number of draws needed to ensure stable results as Halton sequences cover the desired distribution more evenly (Train, 2009).⁴

2.6 Common Specifications in the Literature

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

As working hours are typically concentrated in only few hours categories, most authors include fixed costs of working, hours restrictions, or both in their models. Fixed costs and hours restrictions can also be loosely interpreted as measures for the availability of the respective choice alternatives (Aaberge et al., 2009). Less than half of the models explicitly allow for stigma effects and non-take-up of welfare benefits. This is interesting due to the common finding that the actual benefit participation rate deviates substantially from full take-up. Thus, models that do not explicitly account for the potential disutility are expected to over-predict the number of recipients.

Less variation can be found in terms of the model's treatment of wages. While most studies estimate wages and the labor supply decision separately 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 estimation. In turn, these models neglect potential sample selection issues when estimating wages. There is no consensus in the literature whether predicted wages

⁴ Details on the estimation procedure can be found in Löffler (2013).

Table 1: Model Specifications

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

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

\dagger Robustness checks and alternative model specifications are separated by slashes.

Table 2: Wage Imputation Methods

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

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

should be used only for individuals whose wages are unobserved or for the full sample. Regarding the handling of the wage prediction errors, it becomes increasingly common practice to incorporate and integrate out the errors during the estimation.

3 Data

The baseline estimations in this paper are performed on the German Socio-Economic Panel (SOEP), a representative household panel survey for Germany (Wagner et al., 2007). SOEP includes now more than 24,000 individuals in around 11,000 households. We use the 2008 wave of the SOEP, which provides household data from 2008, as well as data on the labor supply behavior and incomes from the preceding year (i.e., the year before the Great Recession). We rely on the tax and transfer system of 2007, focusing our analysis on the working age population and thus excluding individuals younger than 17 or above the retirement age of 65 from our estimations. Our sample is further restricted to those households where at least one decision maker can freely adjust her labor supply. Therefore, we exclude households where all decision makers are self-employed (since it is difficult to measure true hours and wages for those), civil servants⁵ or in the military service. Moreover, our subsample includes some households with more than two adults, which mainly includes adult children living with their parents. We exclude these young adults from the estimation as it is unclear how their consumption and utility are determined (Dagsvik et al., 2011).

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

For the computation of consumption levels for the different choice categories, we rely on IZA’s policy simulation model IZAΨMOD (v3.0.0), which incorporates a very detailed representation of the German tax and benefit system (see Löffler et al., 2014, for a comprehensive documentation). Some of the estimated models would require

⁵ Tenured civil servants cannot freely adjust the weekly working hours. Note that we keep all other public sector employees.

applying the tax and benefit system for every possible wage rate for every household in every step of the numerical likelihood maximization. To avoid this cumbersome procedure, we approximate the tax and benefit system by using a highly flexible second-degree polynomial that transforms monthly gross earnings into disposable income while controlling for a rich set of household characteristics, as well as all available sources of non-labor income. The resulting R^2 shows a very good fit of more than 99 percent for all population subgroups but single women (only 97 percent for them), which confirms that our approximation performs rather well.⁶ The results are very much in line with those taking advantage of the full representation of the tax and transfer system, we are thus confident that the approximation does not affect our findings.

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

4 Meta-Analysis of Labor Supply Models

Robustness checks in the applied labor supply literature usually narrow down to a small deviation in just one of the modeling assumptions (see Tables 1 and 2). Evers et al. (2008) and Bargain and Peichl (2016) perform meta-analyses of labor supply models comparing estimated labor supply elasticities for different countries and explain them mainly by study characteristics. In either case, it is difficult to draw general conclusions on the exact specification of discrete choice models from the reported results. We overcome these difficulties by estimating a large variety of different modeling assumptions in a controlled environment using the same data. The estimation results allow us to determine how sensitive the estimated outcomes are with respect to the specification and the wage imputation procedure used in the model.

4.1 Set Up of the Analysis

For our analysis, we combine frequently used modeling assumptions and estimate all sensible combinations of these specifications. We estimate 3,456 different model specifications for the five distinct population groups (see Section 3), which leads us to 17,280 maximum likelihood estimations. However, the sample of estimation results is reduced because not all models did converge to a global maximum in a reasonable time span. We drop those estimation results from our analysis that did not converge after 100 iterations of Stata's maximum likelihood implementation. Depending on the labor supply group we lose up to six percent of our sample and end up with 16,730 different estimation results.⁷

⁶ We combine the predicted amounts of consumption with a single random draw for each household; otherwise, we would mistakenly reduce the variance in the consumption variable.

⁷ Of course, more complex models take longer to converge. Apart from that, we do not find systematic effects of different types of assumptions on the probability to converge.

Table 3: Estimated Model Combinations

Model Parameter	Option	All (1)	Number of Converged Models				
			Singles		Couples		
			Male (2)	Female (3)	Male (4)	Female (5)	Both (6)
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 Draws	1,584	1,440	1,564	1,429	1,559	1,456
	5 Draws	1,584	1,571	1,575	1,539	1,582	1,582
Observed Heterogeneity	—	864	835	864	822	860	834
	In β_C Only	864	827	862	834	861	822
	In β_L Only	864	827	858	798	859	836
	In β_L, C	864	810	843	797	849	832
Unobserved Heterogeneity	—	576	574	571	566	570	574
	In β_C Only	864	863	853	846	862	863
	In β_L Only	576	520	574	523	569	541
	In β_L, β_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

Notes: This table shows the number of estimated models over the different model parameters and population subgroups. Column (1) shows the number of possible model combinations for each choice of parameters. Columns (2)-(6) report the number of converged models by population subgroup. Column (2) refers to single male households, column (3) to households with a single female adult (both also including lone parents). Columns (4)-(6) refer to couple households where only the male partner is flexible in his labor supply behavior, where only the female partner is flexible, or where both partners are flexible in their labor supply, respectively.

Table 3 shows the different specifications and the number of converged estimation results. The table reads as follows. We estimate 1,152 distinct models with a Box-Cox transformed utility specification for each of the five labor supply groups. Because few models did not converge to a global maximum in a reasonable amount of time, only 1,022 estimation results for single males and 1,132 for single females are included in our sample. Regardless of the functional form of the utility function, 1,152 of the estimated models neglect any kind of hours restrictions or fixed costs, 1,152 models include part-time restrictions and 1,152 models account for fixed costs of work.

To make the estimation results comparable across the different labor supply groups, we standardize the statistical fit and the estimated elasticities within population groups. We subsequently pool the data and regress the estimation results on indicators for the

different modeling assumptions (mainly represented as dummy variables). We measure the statistical fit by the Akaike Information Criterion (AIC) of the models. To retrieve (uncompensated) labor supply elasticities, we increase the own-wage rates by ten percent and simulate the labor supply reaction to this wage change.⁸

4.2 Empirical Results

The results of these meta-regressions can be found in Table 4. Coefficients have to be compared to the 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. All outcomes are standardized, i.e., coefficients relate to changes in terms of standard deviations, and thus only large estimates (in absolute values) are also economically interesting. Our results show, e.g., that combining this model set-up with a quadratic utility function instead of a translog specification increases the AIC by 12 percent of a standard deviation in the sample and thereby worsens the statistical fit. We summarize the key findings below.⁹

Goodness of Fit. Although the statistical fit is usually not the main outcome of interest, our results show several interesting patterns for future applications (see Table 4). First, the choice of the utility function does not substantially 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 observed labor supply choices, i.e., the AIC decreases. Estimating random coefficients models that also allow for unobserved heterogeneity yields little value-added in terms of the statistical fit – especially compared to the increased computational burden of the estimation. The results regarding the wage imputation show that these modeling decisions affect the statistical fit of the model substantially. Predicting wages not only for non-workers but for the full sample of workers improves the fit significantly. This mirrors the fact that much of the variation in the data is lost by using predicted instead of actual wages for the full sample when not accounting for errors in the wage rate prediction.

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

⁸ Results are robust to different ways of computing own-wage labor supply elasticities, see below for details.

⁹ The presented standard errors do not account for the (potential) variation in the statistical fit and the simulated elasticities for one specific model when estimating the same model using different samples. Accounting for this uncertainty, e.g., by using bootstrap procedures, would produce larger standard errors than those presented but is computationally infeasible in our context due to the large number of estimated models.

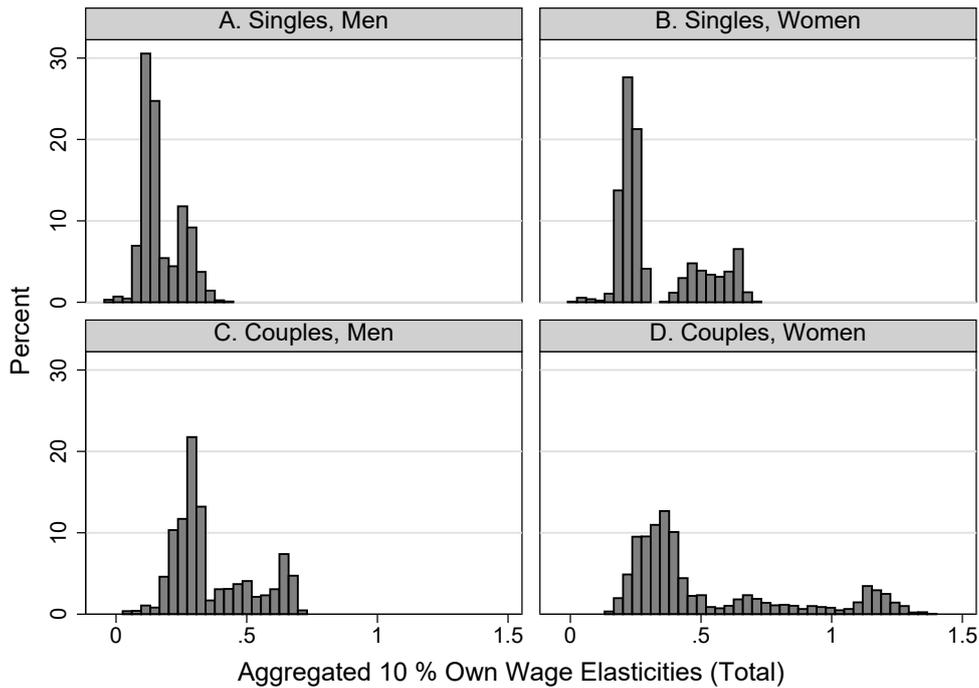
Table 4: Marginal Impact of Modeling Assumptions (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119*** (0.023)	0.124*** (0.028)	-0.015 (0.062)	0.004 (0.053)
Box-Cox	-0.020 (0.026)	0.116*** (0.040)	0.080** (0.035)	0.085** (0.034)
Welfare Stigma	0.968*** (0.076)	0.045 (0.062)	0.065 (0.047)	0.065 (0.042)
Number of Halton Draws	-0.010*** (0.001)	0.005 (0.004)	-0.003 (0.004)	-0.002 (0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647*** (0.082)	0.384*** (0.070)	0.105** (0.039)	0.152*** (0.042)
Fixed Costs	-1.093*** (0.070)	0.481*** (0.067)	0.187*** (0.040)	0.238*** (0.041)
Observed Heterogeneity				
In β_C Only	-0.335*** (0.057)	-0.049 (0.035)	0.060** (0.022)	0.043* (0.023)
In β_L Only	-0.381*** (0.061)	0.048 (0.038)	0.045** (0.021)	0.046* (0.023)
In β_C and β_L	-0.475*** (0.070)	0.016 (0.044)	0.012 (0.019)	0.013 (0.022)
Unobserved Heterogeneity				
In β_C Only	0.005 (0.014)	-0.006 (0.023)	-0.059* (0.032)	-0.051 (0.030)
In β_L Only	0.005 (0.013)	-0.081*** (0.023)	-0.029 (0.028)	-0.037 (0.027)
In β_C And β_L	-0.041*** (0.013)	-0.037 (0.024)	-0.069** (0.027)	-0.064** (0.026)
In β_C And β_L With Correlation	-0.119*** (0.016)	-0.082** (0.034)	-0.102*** (0.034)	-0.101*** (0.033)
Wage Imputation				
Full Sample, No Correction	-0.811*** (0.119)	2.121*** (0.094)	2.235*** (0.091)	2.240*** (0.086)
Full Sample, Error Integrated Out	-0.530*** (0.048)	1.399*** (0.119)	1.385*** (0.123)	1.406*** (0.124)
Full Sample, 1 Random Draw	-0.104** (0.049)	0.071 (0.062)	0.131 (0.093)	0.121 (0.088)
Non-Workers, Error Integrated Out	0.000 (0.067)	0.048 (0.063)	0.040 (0.041)	0.041 (0.041)
Non-Workers, 1 Random Draw	0.070 (0.056)	-0.230*** (0.038)	-0.232*** (0.035)	-0.235*** (0.037)
Constant	1.004*** (0.121)	-0.939*** (0.094)	-0.678*** (0.087)	-0.726*** (0.087)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.849	0.870	0.881

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

Labor Supply Elasticities. More important than the statistical fit is whether specific modeling assumptions systematically influence the out-of-sample predictions when simulating policy or wage changes. Figure 1 shows the distribution of simulated labor supply elasticities across the converged models for four demographic subgroups.¹⁰ The graph shows considerable variation across the different modeling set-ups (within population groups as well as across groups).

Figure 1: Simulated Labor Supply Elasticities For Converged Models



Notes: This figure shows the distribution of estimated labor supply elasticities over the 3,456 different model specifications for four labor supply groups (see panels). Elasticities are defined as hours responses to a ten percent increase in the own wage rate, combining both intensive and extensive margin and aggregating over individual responses. Panel A shows elasticities for single men, Panel B shows the results for single women. We pool estimation results for the three types of couple households and plot the response of the male and female partner in Panel C and Panel D, respectively (see Section 3 for a discussion of the different labor supply types).

In line with the literature, we find that the simulated elasticities are rather robust regarding 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 or full-time jobs. It is thus more likely that people switch from non-participation to 20 or 40 than to 5 or 10 hours of work when accounting for this fact, which leads to higher elasticities at the

¹⁰ We aggregated couples with one and couples with two flexible partners in this figure.

extensive margin.

Substantially more of the variation in simulated elasticities can be explained when analyzing the impact of the wage imputation and the handling of wage prediction errors. Our results thus hold the important message that this part of the model specification is much more relevant to the estimated elasticities than the utility specification. For instance, using predicted wages not only for non-workers but for the full sample of individuals roughly doubles the estimated elasticities. The average own-wage elasticity in our meta-analysis increases from 0.23 to 0.46. This substantial difference can be explained by the fact that predicting wages for the full sample reduces the variance of the wage distribution substantially. To explain the observed working hours with less variation in wages and thus income and consumption, the implied elasticities have to increase. Accounting for wage prediction errors and integrating them out during the estimation markedly reduces the difference. Predicting wages for all individuals but ignoring the wage prediction error yields an average elasticity of 0.65 in our meta-analysis sample as opposed to 0.35 when accounting for the prediction error. Interestingly, the results differ substantially depending on whether a single random draw or higher numbers are used. The *ad hoc* procedure of adding a single random draw tends to cancel out the effect of a full sample prediction, estimated elasticities are close to those of the reference model relying mostly on observed wages (average elasticity of 0.26). In contrast, correcting for the wage prediction error tends to reduce the elasticities, but we still observe the estimated elasticities to be significantly higher than those in which the wage rates were only imputed for non-workers (0.47 vs. 0.25).

Robustness. We perform a wide range of robustness checks to confirm that our results are not driven by the used data or the meta-analysis set up. In particular, we also use a different wave from the same data set and perform our analysis also using data from the CPS for the US (see Table A.1 in the Appendix). In addition to the marginal impact (holding all other specification details constant), we investigate the partial impact of the modeling assumptions (see Table A.2 in the Appendix), which only shows the differences in means due to the specific assumptions (e.g., the mean of elasticities using a translog utility specification vs. the mean of elasticities using different functional forms, irrespective of other modeling issues). The results we obtain are qualitatively the same. We also check the robustness regarding the calculation of elasticities and find no differences whether we simulate one percent or ten percent changes in the own-wage rate (see Table A.3 in the Appendix). Also switching the calculation of the elasticities from aggregated to mean, median or other quantile measures did not affect our findings (see Tables A.4 and A.5 in the Appendix).

Summary. Our results partly confirm previous findings in the literature. While the empirical specification of the systematic utility function has an impact on the statistical fit, we find only little differences in the estimated elasticities. It may thus be justified

to rely on simpler model set-ups when the computational burden is a major concern. However, the majority of the robustness checks applied in the literature focus on the effects of different utility specifications and usually ignore how the underlying wage distribution – and especially the imputation of wages – may influence the results. We find that these assumptions explain much more of the variation in simulated labor supply elasticities than the specification of the utility function. Most previous robustness checks have thus concentrated on issues of secondary order. Instead, more attention should be paid to the wage imputation and the handling of wage prediction errors. Modeling choices regarding the wage handling may thus also explain part of the large variation found in labor supply studies.

Which assumption should be preferred? Integrating out the error term of the wage prediction is clearly preferred over no correction. Thanks to advances in computing power, the additional computational burden should not be an issue anymore. Using only one random draw from the wage distribution, which has been used as a shortcut to avoid long computations, is hence not necessary anymore. In terms of predicting wages for the full sample vs. non-workers only, the answer depends on the research question and data at hand. The first option assumes that all individuals, not only the unemployed, are aware of uncertainties about their individual wage realization and base their labor supply decision on expected wages as derived from the Mincerian wage equation. The second option, on the other hand, assumes that employed workers make choices based on their current wage rate, independent of whether they drew a positive, negative or no wage shock in their current job. Which of these models fits better is a decision that the researcher has to make and it should be made explicit.

5 Conclusion

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

In this paper, we provide an overview of the most important specification issues and conduct a comprehensive sensitivity analysis to disentangle the driving factors behind the results obtained from structural labor supply models. Our results show that even if the modeling assumptions concerning the direct utility specification increase or worsen the statistical fit, i.e., the power to explain the observed labor supply behavior, the models are robust in terms of their implied labor supply elasticities. In contrast, the model predictions are highly sensitive to changes in the underlying wage distribution, a mechanism almost completely neglected in the literature to date. Thus, the questions of whether to use predicted or observed wages for actual workers and whether and how to

integrate the wage prediction error out during the estimation process have a large and statistically significant impact on the statistical fit of the model *and* the estimated labor supply elasticities.

Our findings have important implications for tax policy design. Diamond and Saez (2011) derive simple formulas for the optimal (top) marginal tax rates based upon labor supply elasticities.¹¹ They assume an elasticity of 0.25 as an “a mid-range estimate from the empirical literature” which is close to our mean estimate for models using the observed wage distribution. This leads to an optimal top marginal tax rate of $\tau = \frac{1}{1+1.5 \cdot 0.25} = 72.7$ percent. However, an elasticity of 0.65 as found in models using predicted wages reduces the optimal tax rate to 50.6 percent bringing it closer to actually observed values (the top labor tax rate in the US is 42.5 percent). While we cannot claim that we have identified the true value for the labor supply elasticity – which might not even exist – our analysis shows that more attention should be paid to the specification of structural labor supply models when using them for policy analysis. Future research should try estimating preferences and wages jointly.

¹¹ The formula for the optimal top tax rate is $\tau = \frac{1-g}{1-g+a \cdot e}$ where g is the marginal social welfare weight for the top earners, a is the parameter of the Pareto (income) distribution and e is the labor supply elasticity. Diamond and Saez (2011) assume $g = 0$ to derive the optimal revenue maximizing top tax rate and use an estimated Pareto coefficient of $a = 1.5$ for the US.

References

- Aaberge, R., J. K. Dagsvik, and S. Strøm (1995). "Labor Supply Responses and Welfare Effects of Tax Reforms". *Scandinavian Journal of Economics* 97.4, pp. 635–659.
- Aaberge, R., U. Colombino, and S. Strøm (1999). "Labour Supply in Italy: An Empirical Analysis of Joint Household Decisions, with Taxes and Quantity Constraints". *Journal of Applied Econometrics* 14.4, pp. 403–422.
- Aaberge, R., U. Colombino, and T. Wennemo (2009). "Evaluating Alternative Representations of the Choice Sets in Models of Labor Supply". *Journal of Economic Surveys* 23.3, pp. 586–612.
- Bargain, O. and A. Peichl (2016). "Own-Wage Labor Supply Elasticities: Variation across Time and Estimation Methods". *IZA Journal of Labor Economics* 5.10, pp. 1–31.
- Bargain, O., K. Orsini, and A. Peichl (2014). "Comparing Labor Supply Elasticities in Europe and the US: New Results". *Journal of Human Resources* 49.3, pp. 723–838.
- Bloemen, H. G. and A. Kapteyn (2008). "The Estimation of Utility-Consistent Labor Supply Models By Means of Simulated Scores". *Journal of Applied Econometrics* 23.4, pp. 395–422.
- Blundell, R. and T. E. MaCurdy (1999). "Labor Supply: A Review of Alternative Approaches". In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 3. Amsterdam: North-Holland. Chap. 27, pp. 1559–1695.
- Blundell, R. and A. Shephard (2012). "Employment, Hours of Work and the Optimal Taxation of Low-Income Families". *Review of Economic Studies* 79.2, pp. 481–510.
- Blundell, R., A. Duncan, J. McCrae, and C. Meghir (1999). "Evaluating In-Work Benefit Reform: the Working Families Tax Credit in the UK". Mimeo, presented at the Joint Center for Poverty Research Conference, Northwestern University.
- (2000). "The Labour Market Impact of the Working Families' Tax Credit". *Fiscal Studies* 21.1, pp. 75–104.
- Borjas, G. J. (1980). "The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias". *Journal of Human Resources* 15.3, pp. 409–423.
- Breunig, R., D. A. Cobb-Clark, and X. Gong (2008). "Improving the Modelling of Couples' Labour Supply". *Economic Record* 84.267, pp. 466–485.
- Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins". *American Economic Review: Papers and Proceedings* 101.3, pp. 471–475.
- Dagsvik, J. K., Z. Jia, T. Kornstadt, and T. O. Thoresen (2014). "Theoretical and Practical Arguments For Modeling Labor Supply as a Choice Among Latent Jobs". *Journal of Economic Surveys* 28.1, pp. 134–151.
- Dagsvik, J. K. and S. Strøm (2004). "Sectoral Labor Supply, Choice Restrictions and Functional Form". *Statistics Norway Discussion Paper* 388.
- (2006). "Sectoral Labor Supply, Choice Restrictions and Functional Form". *Journal of Applied Econometrics* 21.6, pp. 803–826.

- Dagsvik, J. K., Z. Jia, K. Orsini, and G. van Camp (2011). "Subsidies on Low-Skilled Workers' Social Security Contributions: The Case of Belgium". *Empirical Economics* 40.3, pp. 779–806.
- Diamond, P. and E. Saez (2011). "The Case for a Progressive Tax: From Basic Research to Policy Recommendations". *Journal of Economic Perspectives* 25.4, pp. 165–190.
- Eissa, N., H. J. Kleven, and C. T. Kreiner (2008). "Evaluation of Four Tax Reforms in The United States: Labor Supply and Welfare Effects For Single Mothers". *Journal of Public Economics* 92.3-4, pp. 795–816.
- Eklöf, M. and H. Sacklén (2000). "The Hausman-MaCurdy Controversy: Why Do the Results Differ across Studies?" *Journal of Human Resources* 35.1, pp. 204–220.
- Ericson, P. and L. Flood (1997). "A Monte Carlo Evaluation of Labor Supply Models". *Empirical Economics* 22.3, pp. 431–460.
- Euwals, R. and A. van Soest (1999). "Desired and Actual Labour Supply of Unmarried Men and Women in the Netherlands". *Labour Economics* 6.1, pp. 95–118.
- Evers, M., R. de Mooij, and D. van Vuuren (2008). "The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates". *De Economist* 156.1, pp. 25–43.
- Flood, L., J. Hansen, and R. Wahlberg (2004). "Household Labor Supply and Welfare Participation in Sweden". *Journal of Human Resources* 39.4, pp. 1008–1032.
- Flood, L., R. Wahlberg, and E. Pylkkänen (2007). "From Welfare to Work: Evaluating a Tax and Benefit Reform Targeted at Single Mothers in Sweden". *LABOUR* 21.3, pp. 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, pp. 251–256.
- Hausman, J. A. (1981). "Labor Supply". In: *How Taxes Affect Economic Behavior*. Ed. by H. J. Aaron and J. A. Pechman. Washington, D.C.: Brookings Institution, pp. 27–72.
- Heckman, J. J. (1993). "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review: Papers and Proceedings* 83.2, pp. 116–121.
- Hoynes, H. W. (1996). "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP". *Econometrica* 64.2, pp. 295–332.
- Immervoll, H., H. J. Kleven, C. T. Kreiner, and N. Verdellin (2011). "Optimal Tax and Transfer Programs For Couples With Extensive Labor Supply Responses". *Journal of Public Economics* 95.11, pp. 1485–1500.
- Keane, M. P. and R. Rogerson (2012). "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom". *Journal of Economic Literature* 50.2, pp. 464–476.
- Keane, M. P. (2010). "Structural vs. Atheoretic Approaches to Econometrics". *Journal of Econometrics* 156.1, pp. 3–20.
- Keane, M. P. and R. Moffitt (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply". *International Economic Review* 39.3, pp. 553–589.
- Keane, M. P. and N. Wasi (2012). "Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior". *Journal of Applied Econometrics* 28.6, pp. 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.
- Löffler, M., A. Peichl, N. Pestel, S. Siegloch, and E. Sommer (2014). "Documentation IZAΨMOD v3.0: The IZA Policy Simulation Model". *IZA Discussion Paper* 8553.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- MaCurdy, T. E., D. Green, and H. Paarsch (1990). "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply". *Journal of Human Resources* 25.3, pp. 412–490.
- Manski, C. F. (2014). "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy". *Quantitative Economics* 5.1, pp. 145–174.
- McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York: Academic Press. Chap. 4, pp. 105–142.
- Moffitt, R. (1983). "An Economic Model of Welfare Stigma". *American Economic Review* 73.5, pp. 1023–1035.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press.
- van Soest, A. (1995). "Structural Models of Family Labor Supply – A Discrete Choice Approach". *Journal of Human Resources* 30.1, pp. 63–88.
- van Soest, A. and M. Das (2001). "Family Labor Supply and Proposed Tax Reforms in the Netherlands". *De Economist* 149.2, pp. 191–218.
- van Soest, A., M. Das, and X. Gong (2002). "A Structural Labour Supply Model With Flexible Preferences". *Journal of Econometrics* 107.1-2, pp. 345–374.
- Wagner, G. G., J. R. Frick, and J. Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". *Journal of Applied Social Science Studies* 127.1, pp. 139–169.
- Ziliak, J. P. and T. J. Kniesner (1999). "Estimating Life-Cycle Labor Supply Tax Effects". *Journal of Political Economy* 107.2, pp. 326–359.

A Online Appendix

Table A.1: Marginal Impact of Modeling Assumptions (CPS)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.640*** (0.062)	0.217 (0.183)	0.207 (0.185)	0.210 (0.185)
Number of Halton Draws	-0.015*** (0.002)	0.022** (0.008)	0.023** (0.008)	0.022** (0.008)
Hours Restrictions				
Part-Time Restrictions	-1.855*** (0.089)	0.420** (0.188)	0.397* (0.189)	0.403* (0.190)
Fixed Costs	-1.279*** (0.067)	0.192 (0.120)	0.125 (0.104)	0.142 (0.106)
Observed Heterogeneity				
In β_C Only	-0.138*** (0.015)	-0.152*** (0.036)	-0.051 (0.031)	-0.078** (0.031)
In β_L Only	-0.258*** (0.026)	-0.066* (0.036)	-0.080* (0.044)	-0.076* (0.042)
In β_C And β_L	-0.309*** (0.027)	-0.115** (0.040)	-0.094* (0.044)	-0.097** (0.043)
Unobserved heterogeneity				
In β_C Only	0.067*** (0.013)	-0.118*** (0.037)	-0.113** (0.044)	-0.114** (0.043)
In β_L Only	0.070*** (0.007)	-0.119** (0.053)	-0.121** (0.054)	-0.120** (0.055)
In β_C And β_L	0.046*** (0.009)	-0.089** (0.032)	-0.083** (0.036)	-0.084** (0.035)
In β_C And β_L With Correlation	0.021*** (0.007)	-0.063** (0.029)	-0.058** (0.025)	-0.059** (0.026)
Wage Imputation				
Full Sample, No Correction	-0.111* (0.057)	0.912*** (0.284)	0.918*** (0.300)	0.921*** (0.299)
Full Sample, 1 Random Draw	0.025 (0.046)	0.338 (0.244)	0.428* (0.229)	0.413* (0.230)
Non-Workers, 1 Random Draw	0.030 (0.053)	-0.329 (0.362)	-0.237 (0.362)	-0.255 (0.363)
Constant	0.832*** (0.098)	-0.813*** (0.271)	-0.864*** (0.270)	-0.857*** (0.270)
Labor Supply Types Fixed Effects	Yes	Yes	Yes	Yes
Observations	4,305	3,439	3,439	3,439
Explanatory Power R^2	0.820	0.353	0.340	0.344

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

Table A.2: Partial Impact of Modeling Assumptions (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Translog	-0.045* (0.024)	-0.125*** (0.021)	-0.035 (0.045)	-0.047 (0.040)
Quadratic	0.135*** (0.013)	0.067* (0.039)	-0.054 (0.053)	-0.037 (0.046)
Box-Cox	-0.093*** (0.017)	0.061 (0.049)	0.094** (0.034)	0.090** (0.034)
Welfare Stigma	0.965*** (0.076)	0.051 (0.061)	0.072 (0.047)	0.071 (0.042)
Number of Halton Draws	-0.013*** (0.003)	0.008 (0.007)	-0.003 (0.007)	-0.001 (0.007)
Hours Restrictions				
None	1.376*** (0.075)	-0.425*** (0.067)	-0.139*** (0.038)	-0.188*** (0.039)
Part-Time Restrictions	-1.110*** (0.052)	0.145*** (0.041)	0.013 (0.024)	0.035 (0.026)
Fixed Costs	-0.244*** (0.034)	0.278*** (0.033)	0.127*** (0.024)	0.153*** (0.023)
Observed Heterogeneity				
None	0.398*** (0.063)	-0.002 (0.038)	-0.035* (0.019)	-0.030 (0.021)
In β_C Only	-0.046** (0.017)	-0.070*** (0.015)	0.042*** (0.014)	0.024* (0.013)
In β_L Only	-0.121*** (0.020)	0.067*** (0.015)	0.028* (0.014)	0.035** (0.014)
In β_C And β_L	-0.235*** (0.031)	0.004 (0.022)	-0.036*** (0.010)	-0.030*** (0.010)
Unobserved Heterogeneity				
None	0.057 (0.040)	0.090 (0.110)	0.125 (0.117)	0.122 (0.117)
In β_C Only	0.029* (0.015)	0.075** (0.036)	0.013 (0.038)	0.023 (0.038)
In β_L Only	0.050 (0.040)	-0.123 (0.110)	-0.032 (0.110)	-0.047 (0.111)
In β_C And β_L	-0.035** (0.015)	0.039 (0.039)	0.006 (0.038)	0.011 (0.038)
In β_C And β_L With Correlation	-0.102** (0.039)	-0.128 (0.102)	-0.124 (0.111)	-0.127 (0.110)
Wage Imputation				
Full Sample Imputation	-0.498*** (0.100)	1.248*** (0.288)	1.313*** (0.294)	1.317*** (0.296)
Error Integrated Out	-0.037 (0.125)	0.267 (0.351)	0.190 (0.359)	0.207 (0.362)
Full Sample, No Correction	-0.720*** (0.119)	1.921*** (0.145)	2.033*** (0.144)	2.036*** (0.142)
Full Sample, Error Integrated Out	-0.334*** (0.081)	1.004*** (0.239)	0.935*** (0.253)	0.960*** (0.254)
Full Sample, 1 Random Draw	0.143 (0.089)	-0.599** (0.237)	-0.554** (0.258)	-0.569** (0.257)
Non-Workers, Error Integrated Out	0.269*** (0.094)	-0.544** (0.227)	-0.606** (0.230)	-0.602** (0.231)
Observations	16,730	13,219	13,219	13,219

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

Table A.3: Marginal Impact, Aggregated 1 % Elasticities (SOEP)

	Fit	1 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119*** (0.023)	0.132*** (0.028)	0.030 (0.050)	0.043 (0.043)
Box-Cox	-0.020 (0.026)	0.133*** (0.042)	0.087** (0.035)	0.094** (0.035)
Welfare Stigma	0.968*** (0.076)	-0.028 (0.084)	0.084 (0.053)	0.071 (0.044)
Number of Halton Draws	-0.010*** (0.001)	0.007* (0.004)	-0.002 (0.004)	-0.001 (0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647*** (0.082)	0.390*** (0.071)	0.134*** (0.038)	0.176*** (0.041)
Fixed Costs	-1.093*** (0.070)	0.494*** (0.068)	0.217*** (0.039)	0.264*** (0.041)
Observed Heterogeneity				
In β_C Only	-0.335*** (0.057)	-0.057 (0.035)	0.060** (0.022)	0.042* (0.022)
In β_L Only	-0.381*** (0.061)	0.032 (0.037)	0.041* (0.022)	0.041* (0.023)
In β_C And β_L	-0.475*** (0.070)	-0.002 (0.044)	0.016 (0.020)	0.013 (0.022)
Unobserved Heterogeneity				
In β_C Only	0.005 (0.014)	-0.009 (0.024)	-0.054* (0.031)	-0.048 (0.030)
In β_L Only	0.005 (0.013)	-0.085*** (0.024)	-0.032 (0.028)	-0.040 (0.027)
In β_C And β_L	-0.041*** (0.013)	-0.036 (0.025)	-0.068** (0.027)	-0.064** (0.026)
In β_C And β_L With Correlation	-0.119*** (0.016)	-0.087** (0.033)	-0.092** (0.033)	-0.093*** (0.033)
Wage Imputation				
Full Sample, No Correction	-0.811*** (0.119)	2.089*** (0.117)	2.245*** (0.088)	2.248*** (0.084)
Full Sample, Error Integrated Out	-0.530*** (0.048)	1.427*** (0.103)	1.398*** (0.114)	1.425*** (0.117)
Full Sample, 1 Random Draw	-0.104** (0.049)	0.086 (0.085)	0.102 (0.078)	0.100 (0.079)
Non-Workers, Error Integrated Out	0.000 (0.067)	0.054 (0.062)	0.046 (0.035)	0.048 (0.038)
Non-Workers, 1 Random Draw	0.070 (0.056)	-0.157*** (0.056)	-0.220*** (0.028)	-0.214*** (0.032)
Constant	1.004*** (0.121)	-0.930*** (0.104)	-0.730*** (0.084)	-0.770*** (0.087)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.816	0.880	0.889

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

Table A.4: Marginal Impact, Mean 10 % Elasticities (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119*** (0.023)	0.100*** (0.033)	0.022 (0.047)	0.041 (0.038)
Box-Cox	-0.020 (0.026)	0.101** (0.041)	0.097*** (0.034)	0.096*** (0.033)
Welfare Stigma	0.968*** (0.076)	-0.026 (0.063)	0.034 (0.045)	0.026 (0.039)
Number of Halton Draws	-0.010*** (0.001)	0.005 (0.004)	-0.003 (0.004)	-0.001 (0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647*** (0.082)	0.353*** (0.089)	0.169*** (0.045)	0.219*** (0.056)
Fixed Costs	-1.093*** (0.070)	0.448*** (0.087)	0.254*** (0.044)	0.307*** (0.053)
Observed Heterogeneity				
In β_C Only	-0.335*** (0.057)	0.037 (0.036)	0.048* (0.024)	0.046 (0.027)
In β_L Only	-0.381*** (0.061)	0.187*** (0.040)	0.068*** (0.022)	0.101*** (0.027)
In β_C And β_L	-0.475*** (0.070)	0.187*** (0.053)	0.036 (0.022)	0.074** (0.029)
Unobserved Heterogeneity				
In β_C Only	0.005 (0.014)	0.001 (0.022)	-0.046 (0.031)	-0.036 (0.029)
In β_L Only	0.005 (0.013)	-0.075*** (0.026)	-0.031 (0.028)	-0.041 (0.026)
In β_C And β_L	-0.041*** (0.013)	-0.027 (0.025)	-0.059** (0.026)	-0.053** (0.025)
In β_C And β_L With Correlation	-0.119*** (0.016)	-0.083** (0.036)	-0.097*** (0.033)	-0.098*** (0.033)
Wage Imputation				
Full Sample, No Correction	-0.811*** (0.119)	2.130*** (0.106)	2.264*** (0.092)	2.267*** (0.089)
Full Sample, Error Integrated Out	-0.530*** (0.048)	1.265*** (0.134)	1.365*** (0.132)	1.364*** (0.140)
Full Sample, 1 Random Draw	-0.104** (0.049)	0.058 (0.049)	0.146 (0.087)	0.122 (0.078)
Non-Workers, Error Integrated Out	0.000 (0.067)	0.035 (0.053)	0.062 (0.042)	0.057 (0.043)
Non-Workers, 1 Random Draw	0.070 (0.056)	-0.164*** (0.028)	-0.220*** (0.034)	-0.210*** (0.035)
Constant	1.004*** (0.121)	-0.966*** (0.119)	-0.749*** (0.091)	-0.814*** (0.097)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.820	0.876	0.883

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

Table A.5: Marginal Impact, Median 10 % Elasticities (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119*** (0.023)	0.107*** (0.030)	0.079 (0.063)	0.098* (0.048)
Box-Cox	-0.020 (0.026)	0.084** (0.036)	0.042 (0.044)	0.056 (0.040)
Welfare Stigma	0.968*** (0.076)	0.001 (0.059)	0.062 (0.068)	0.059 (0.062)
Number of Halton Draws	-0.010*** (0.001)	0.005 (0.005)	-0.005 (0.005)	-0.003 (0.005)
Hours Restrictions				
Part-Time Restrictions	-1.647*** (0.082)	0.370*** (0.084)	0.046 (0.069)	0.116* (0.058)
Fixed Costs	-1.093*** (0.070)	0.448*** (0.085)	0.112 (0.076)	0.181*** (0.063)
Observed Heterogeneity				
In β_C Only	-0.335*** (0.057)	0.046 (0.035)	0.010 (0.023)	0.013 (0.022)
In β_L Only	-0.381*** (0.061)	0.136*** (0.037)	-0.034 (0.022)	-0.001 (0.023)
In β_C And β_L	-0.475*** (0.070)	0.130*** (0.044)	-0.103*** (0.025)	-0.068*** (0.024)
Unobserved Heterogeneity				
In β_C Only	0.005 (0.014)	-0.020 (0.027)	-0.098** (0.038)	-0.088** (0.036)
In β_L Only	0.005 (0.013)	-0.077*** (0.026)	-0.010 (0.038)	-0.017 (0.036)
In β_C And β_L	-0.041*** (0.013)	-0.052** (0.025)	-0.096*** (0.030)	-0.088*** (0.029)
In β_C And β_L With Correlation	-0.119*** (0.016)	-0.107** (0.040)	-0.132*** (0.036)	-0.122*** (0.036)
Wage Imputation				
Full Sample, No Correction	-0.811*** (0.119)	2.169*** (0.098)	2.052*** (0.123)	2.110*** (0.112)
Full Sample, Error Integrated Out	-0.530*** (0.048)	1.278*** (0.140)	1.390*** (0.106)	1.397*** (0.103)
Full Sample, 1 Random Draw	-0.104** (0.049)	0.060 (0.053)	0.227 (0.140)	0.199 (0.127)
Non-Workers, Error Integrated Out	0.000 (0.067)	0.042 (0.050)	0.013 (0.035)	0.021 (0.032)
Non-Workers, 1 Random Draw	0.070 (0.056)	-0.187*** (0.033)	-0.265*** (0.047)	-0.266*** (0.042)
Constant	1.004*** (0.121)	-0.940*** (0.114)	-0.529*** (0.104)	-0.623*** (0.096)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.832	0.769	0.806

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