CESifo WORKING PAPERS

10827
2023

December 2023

Evaluating the Discrete Choice and BN Methods to Estimate Labor Supply Functions

Sören Blomquist

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10827*

Evaluating the Discrete Choice and BN Methods to Estimate Labor Supply Functions

Abstract

Estimated labor supply functions are important tools when designing an optimal income tax or calculating the effect of tax reforms. It is therefore of large importance to use estimation methods that give reliable results and to know their properties. In this paper Monte Carlo simulations are used to evaluate two different methods to estimate labor supply functions; the discrete choice method and a nonparametric method suggested in Blomquist and Newey (2002). The focus is on the estimators' ability to predict the hours of work for a given tax system and the change in hours of work when there is a tax reform. The simulations show that the DC method is quite sensitive to misspecifications of the likelihood function and to measurement errors in hours of work. A version of the Blomquist Newey method shows the overall best performance to predict the hours of work.

JEL-Codes: C400, C520, C530, H200, H300.

Keywords: labor supply, tax reform, predictive power, estimation methods, Monte Carlo simulations.

Sören Blomquist
Uppsala Center for Fiscal Studies at the
Department of Economics
Uppsala University / Sweden
soren.blomquist@nek.uu.se

2023-12-04
The author wants to thank Max Löffler for comments and Anil Kumar and Whitney Newey for discussions. Handelsbankens forskningsstiftelser provided financial support.

1. Introduction.

The labor supply and taxable income elasticities are key parameters when designing an optimal income tax, and ministries of finance around the world use estimates from labor supply and taxable income functions to calculate the effects of various types of tax reform. It is therefore of large importance to use estimation methods that give reliable results and to know their properties. The present paper studies the Discrete Choice (DC) maximum likelihood method to estimate how labor supply reacts to changes in the tax system and compares it with a method to estimate expected hours of work developed in Blomquist and Newey (2002), the BN method.

As other maximum likelihood methods, the DC-method assumes a specific parametric data generating process (dgp). For real world data we will never know the actual data generating process. Hence, the DC-method likelihood function applied to real data will always be misspecified. It is therefore of interest to know how the DC-method holds up under conditions such that the DC likelihood function is misspecified. The BN-method builds on less specific assumptions about the data generating process. The basic assumption is that data is generated by utility maximization with convex preferences subject to a convex budget constraint. The method is nonparametric and allows for optimization/measurement errors and heterogenous preferences. We expect the DC-method to perform better than the BN-method when the dgp we use to generate data fully aligns with the DC likelihood function. After all, if the likelihood function is correctly specified the maximum likelihood estimator is an efficient estimator. However, it is possible that the BN-method is more robust and performs well under a wider set of dgp:s than the DC-method. Monte Carlo simulations will be used to see to what extent this is true.

The DC method[2]

McFadden (1974) developed a discrete choice model of how individuals choose transportation mode to get to work. McFadden's model has been adapted so it can be used to estimate a model of how hours of work are chosen. According to Aaberge and Colombino (2014) the DC method is the dominating method to calculate the effect of tax reform on hours of work. This method was introduced by Dagsvik (1994), Aaberge et al. (1995) and van Soest (1995). Blundell and Shephard (2012) and Beffy et al. (2019) are recent studies using the method. It is used by the Swedish ministry of finance and many other ministries of finance around the world to make predictions of the effect of tax reform. The method allows for co-variates, but I have not seen any description in the DC literature of how to handle endogeneity of budget set variables.

In the discrete choice method applied to hours of work, the budget set is represented by a discrete number of points and, since measured hours of work is a continuous variable, a rule how to map the measured hours into the set of discrete points. Individuals are assumed to maximize utility over the fixed number of points. Given this description of how hours of work are chosen, the Conditional Multinomial logit model can be used for estimation. In the McFadden model, there is a particular form of heterogeneity in individuals' preferences, but there is no random term representing optimization and/or measurement errors in the dependent

[2] The DC method to estimate a labor supply model is described in more detail in section 2 and Appendix A.

variable. That might not be a problem when it comes to the choice between modes of transportation, but might be a problem when the DC method is used to estimate an hours of work model. That the DC method does not allow for optimization/measurement errors in the dependent variable sets the DC method apart from regression models to estimate labor supply functions; regression models do account for errors in the dependent variable.

In the original literature on discrete choice, a measurement error in the dependent variable means that observations are misclassified. Meyer and Mittag (2017) study misclassification in a binary choice model and find that misclassification leads to biased and inconsistent estimates. This result leads one to suspect that measurement error in hours of work might yield biased and inconsistent estimates, when the multinomial logit estimator is applied to hours of work data that has been converted into a form so they fit the discrete choice framework.

The BN method[3]

The BN method, developed in Blomquist and Newey (2002), is a nonparametric method to estimate an hours of work function. The individuals' choice of hours of work is seen as a function of the entire budget set. The method takes explicit account of preference heterogeneity and optimization/measurement errors. The estimated function can be used to predict the expected hours of work for a given budget constraint, where the expectation is taken over a heterogeneous multidimensional preference distribution. A convex, continuous, piece-wise linear budget set, is exactly described by the slopes and intercepts of the linear segments. If the budget constraint consists of many segments the dimensionality becomes a problem. However, Blomquist and Newey (2002) show that if the hours of work is determined by utility maximization with convex preferences and convex budget constraints, then one can get an object with a lower dimension to estimate. The function is estimated by series estimation; power series of the slopes and intercepts are used to approximate the hours of work function. One of the challenges is to choose terms to be included in the regression. I will use the Lasso method, introduced by Tibshirani (1996), to choose the order in which to include variables in the regression and a cross-validation measure to decide how many regressors to include. In the tables I will show predictions using the Lasso-coefficients, but, following the suggestion in Belloni and Chernozhukow (2013), I will also show predictions using OLS on the set of regressors chosen by Lasso (post-Lasso). The BN method allows for additive as well as multiplicative optimization/measurement errors in hours of work. The method allows for co-variates and can handle endogeneity of budget set variables, if instruments are available.

I will use Monte Carlo simulations to study the properties of the two methods. Section 2 gives a brief description of the Monte Carlo simulations and the shape of the budget constraint. In section 3 I describe results where data are generated from a discrete dgp and in section 4 results where data are generated from a continuous dgp. The utility functions and tax systems for which hours of work will be predicted are such that the average annual hours of work are around 2000 and decrease when taxes change from tax system 4 to tax system 5. Section 5 gives a brief summary.

[3] The BN-method is described in greater detail in Appendix B

2. Monte Carlo simulations

I will use Monte Carlo simulations to study the properties of the DC and BN methods under a wide range of dgp:s, some with continuous budget constraints and some with a number of discrete points. The simulations are meant to mimic the situation of a policy maker that has data from four different tax systems at different points in time. The policy maker uses the data to estimate an hours of work function, which is used to predict the effect of a tax reform. The setup of the simulations is such that it illustrates the methods ability to predict how the *intensive* choice of hours of work is affected by taxes. Since the neglect to account for optimization /measurement errors for the dependent variable sets the DC method apart from the BN method and other regression methods to estimate labor supply, a subset of the simulations will study the effect of such errors. However, I will also study the effect of heterogeneity in preferences, quantity constraints and the effect of sample size. The simulations will mostly be for convex budget constraints, but I will also study how the estimators perform when the budget constraints are nonconvex. I will leave for future research to study measurement errors in independent variables. Blomquist (1996) show that some types of measurement errors in the independent variables can cause serious bias. If instrumental variables are available the BN method can use control functions to handle this problem. Measurement errors in the independent variables are rarely discussed in the DC literature, although Löffler et.al (2018) gives a detailed account of the sensitivity of estimates depending on how wage rates are measured and represented in the estimation.

Taxes and Construction of the Budget Constraint

Let w be the hourly wage rate, h hours of work, c consumption, y_j the virtual income (intercept) for the $j:th$ segment, cap_{nt} nontaxable capital income and cap_t taxable capital income. We will consider a tax system where taxable capital income and labor income are taxed jointly according to the tax function $T\left(cap_t + wh\right)$. The budget constraint can then be written as: $c = cap_{nt} + cap_t + wh - T(cap_t + wh)$. Since real world tax systems are usually piece wise linear, we consider such systems. For such a system capital income and labor income are taxed with the same rate for a given linear segment. The effect of the joint taxation of capital income and labor income is to affect the kink points in terms of hour of work. Let the kink points in terms of taxable income be denoted by A_j, and kink points in terms of hours of work l_j, with A_j and l_j denoting the left endpoint of the $j:th$ segment and with $A_1 = l_1 = 0$. The kink points for a tax system is given in taxable income, the kink points in terms of hours of work, l_j, are then defined by the relation $\tilde{l}_j w + cap_t = A_j$, or $\tilde{l}_j = \left(A_j - cap_t\right)/w$, which gives $l_j = \max\left(0, \tilde{l}_j\right)$, since hours of work cannot be negative. Within a given segment j, defined by the kink points l_j, l_{j+1}, the tax rate is t_j and the slope of the budget constraint $w_j = w\left(1 - t_j\right)$. The intercept for the first segment is given by $y_1 = cap_{nt} + cap_t - T\left(cap_t + 0\right)$. The intercepts for the other segments are given by the recursive formula $y_j + w\left(1-t_j\right)l_j = y_{j-1} + w\left(1-t_{j-1}\right)l_j$. We rewrite this as

$$y_j = y_{j-1} + wl_j\left(t_j - t_{j-1}\right) = \quad y_{j-1} + \frac{w\left(A_j - cap_t\right)}{w}\left(t_j - t_{j-1}\right) = y_{j-1} + \left(A_j - cap_t\right)\left(t_j - t_{j-1}\right)$$

It is interesting to note that if there were no taxable capital income, the virtual incomes would be the same for all individuals facing the same tax system, except for a vertical shift due to different nontaxable capital incomes. Also, the difference $y_j - y_{j-1} = \left(A_{j-1} + cap_t\right)\left(t_j - t_{j-1}\right)$ would be the same for everyone facing the same tax system, if there were no taxable capital income. However, if there is taxable capital income there will also be variation across individuals facing the same tax system.

The five tax systems considered will all be piecewise linear and consist of 4 linear segments. The data generated for the first four tax systems will be used to estimate labor supply functions using the DC and BN methods. These functions are then used to predict average hours of work for tax systems 4 and 5 and also the change in hours of work when there is a tax reform that changes taxes from tax system 4 to tax system 5. The kink points for tax systems 4 and 5 are the same, in terms of thousands of kronor they are $k = [0 \;\; 80 \;\; 160 \;\; 250]$. The marginal tax rates for ax system 4 are $m = [0 \;\; 0.2 \;\; 0.3 \;\; 0.4]$ and for tax system 5 $m = [0 \;\; 0.25 \;\; 0.35 \;\; 0.45]$, so for the segments where there is a positive marginal tax rate the marginal tax is increased by 5 percentage points.

In the Monte Carlo simulations I will use 10,000 observations for each period, implying that for each replication of the model 40,000 observation will be used for the estimation. There will be 100 replications of the model.

3. Simulations where individuals choose from a discrete set of points

MacFadden's discrete choice model describes how individuals choose transportation mode to get to work. Suppose there are *N* individuals and that each individual can choose between *J* alternative transportation modes. Each alternative *j* gives a utility

$$v\left(x_j^i\right) = u\left(x_j^i\right) + \varepsilon_j^i, \; j = 1,..,J, \; i=1,..,N \qquad (1)$$

where $u\left(x_j^i\right)$ is a quasi-concave function of a vector x_j,which shows relevant characteristics of the different travel alternatives, such as cost of transportation, travel time and risk of accident. Each individual chooses the alternative that gives the highest utility. Assuming the ε_j^i are drawings from the GEV 1 extreme value distribution, the likelihood function takes a convenient form and the parameters of the utility function can be estimated by the maximum likelihood method. It is possible to think of convex combinations of the characteristics x_j and that convex combinations are weakly preferred compared to the baskets from which the convex combinations are constructed. It is also reasonable to think of aspects of the alternatives where it is less reasonable to think of convex combinations. What is the convex combination of a train and a bus? Still, some people might enjoy riding a bus; others might think it is fun to ride on a train etc. These aspects can be captured by the random terms ε_j^i in equation (1). It should be noted that, although $u(x_j)$ is quasi-concave, the function $v(x_j)$ is not.

The DC model to choose hours of work follows the original MacFadden model closely. Suppose there are N individuals and that each individual can choose between J alternative values for hours of work. Each alternative j gives a utility described by

$$v(c_{j,}^i h_j^i) = u(c_{j,}^i h_j^i) + \varepsilon_j^i, \; j = 1,..,J, \; i=1,..,N \qquad (2)$$

where $u(c_{j,}^i h_j^i)$ is a quasi-concave function of consumption, c, and hours of work, h, associated with the $j:th$ alternative and ε_j^i is a term drawn from the GEV 1 extreme value distribution.

Although the MacFadden model seems well suited to describe the choice of mode of transportation, the model might be less adequate to describe the choice of hours of work. The part with a quasi-concave utility function is well in line with how preferences usually are modelled in economics. However, adding the GEV random terms implies that the preferences sometimes will become non-convex. My interpretation of this assumption is that it is not because there is any empirical support for it, but that it is a technical assumption made to motivate the form of the likelihood function. I will present simulations where we have this kind of non-convex preferences in the dgp, but in the majority of the simulations we will not add GEV errors to the utilities given by $u(c_{j,}^i h_j^i)$.

The discrete choice estimation literature assumes a particular data generating process (dgp). If we in the simulations use a dgp which completely agrees with the dgp assumed when constructing the likelihood function we would find that, for large samples, the discrete choice method would perform very well. However, that is of little interest and says nothing on how well the method performs on real data. It would just confirm what is since long well known, that the maximum likelihood method yields consistent estimates if the assumptions underlying the construction of the likelihood function are correct. Still, as a starting point we will consider the case where the dgp is exactly the same as the one assumed when constructing the likelihood function. We will then introduce changes in the data generating process and study how the performance of the discrete choice and BN methods are affected.

In the simulations I will assume individuals can choose from 11 combinations of annual hours of work and consumption, the hours of work starting at zero hours and being 300 hours apart (0,300,600,…,3000). In the first set of simulations the set of discrete set of points individuals can choose from in the dgp will be the same set that will be used in the estimation. GEV errors will be added to generate data. The utility function used in the data generating process will also be used in the estimation. Several utility functions have been used in the DC-literature. Löffler et al. (2018) finds that results do not differ much when they try different utility functions in the estimation. Since the translog utility function seems to be one of the more popular utility functions to use, we will start out with this utility function.

Measuring hours of work in thousands of hours and setting the time endowment to 4, we can express the utility function in terms of leisure and consumption; $U(C,L) = u(c, 4-h)$. In the first three simulations I will use the translog utility function to generate data. This utility function is linear in parameters. This implies that if we have a parameter vector β that generates a certain distribution of hours of work, then all parameter vectors $\tilde{\beta} = \gamma\beta$, where γ is a positive constant, will generate exactly the same distribution of hours of work, if there are no GEV errors in the dgp. If there are GEV errors in the dgp, the GEV errors will not affect the distribution of hours of work if the errors are small in comparison with the translog utility values, but will affect the distribution if the GEV errors are large in comparison with the translog utility function.

Data are generated for five different tax systems; we let the wage distribution, the distribution of taxable nonlabor income and nontaxable nonlabor income differ between the data sets, but we use the same distributions for data sets four and five. The only difference between data sets 4 and 5 are the tax systems. We estimate on pooled data from four of the datasets using both the DC method and the BN method. We then predict hours of work for tax systems 4 and 5. The prediction for tax system 4 is a within sample prediction, while the prediction for tax system 5 is an out of sample prediction.

We have no fixed costs of work and a distribution of preferences and wage rate so that the average hours of work are around 2,000 per year and no zero hours of work, so the simulations are meant to see how well the methods can predict the intensive choice of hours of work. In the Monte Carlo simulations I do 100 replications. When we do the simulations the 5 tax system stays constant over the 100 replications of the simulation program, but for each replication we make new drawings of the distributions of wage rates, taxable and nontaxable nonlabor income, and in the relevant simulations random preferences and measurement errors.

In the first three simulations we use the translog utility function:

$$U = \beta_1 L + \beta_2 C + \beta_3 L^2 + \beta_4 C^2 + 2\beta_5 LC \qquad (3)$$

where *L* is the log of leisure time and *C* is the log of consumption. Identification is obtained by stipulating that the random terms in eq. (2) are drawings from the GEV 1 extreme value distribution.

In the computations with the *translog* utility function we will use thousands of hours of work and hundred thousands of "kronor". In this way the hours and consumption will be of the same numerical magnitude, which will facilitate the numerical computations. Therefore the parameters in eq. (3) will be given with the understanding that the input into the function are in thousands of hours of work and thousands of kronor. For increased readability we in the tables and the text will express the labor supply in hours of work.

Tables 1-3 show how results change as we successively increase the importance of the GEV errors. We will use the translog utility function with parameter vectors proportional to β = (175 380 -7 -10 0). As a measure of the relative importance of the $U(C_j, L_j)$ function and the GEV errors I will use the ratio of the mean of the absolute values of the GEV errors, where the mean is taken over 10,000,000 drawings from the GEV distribution, to the absolute value of the utility function evaluated at L=2 leisure hours of work and C=2.2, where leisure is measured in thousands of hours and C in hundred thousands of kronor of consumption; these numbers are typical at individuals' optima. I denote the ratio by ρ. For the simulations shown in table 1 the parameter vector β has been multiplied by a very large number so that the ratio ρ goes to zero. For the simulations shown in table 2 the β vector is multiplied by one. The mean of the absolute values of the GEV errors is 1.0161. The value of the utility function evaluated at L=2 and C=2.2 is 411.33 for the parameter vector β=(175 380 -7 -10 0), so the ratio of the mean of absolute vales of ε to the absolute value of U is 0.0025 for the simulations shown in table 2. For the simulations shown in table 3 we have multiplied the parameter vector $\beta =$ by 0.05, so the parameter vector becomes $\beta =$ (8.75 19 -0.35 -0.5 0) and the ratio ρ becomes 0.05, i.e. 20 times larger than for the results shown in table 2.

I define the true average hours of work for tax systems 4 and 5 as the averages taken over 10 million individuals. In the tables I show the true average hours of work for tax systems 4 and 5 as well as the averages taken over the 100 replications of the predicted values using the DC and BN methods. For the DC method we show results both when GEV errors are added in the prediction (DC1) and when they are not (DC2). For the BN method we show results both for the post Lasso (BN-OLS) and the Lasso (BN-Lasso). We also show the percentage error of the predictions. The change in hours of work can be calculated using either the value of average hours of work for the sample of data from tax system 4 (Measure one) or the predicted value (Measure two). Since the sample with the data from tax system 4 is a random sample from the true data generating process, it gives un unbiased estimate of true average hours of work for the tax system. Below the predictions we give the root mean square error and the standard deviation of the estimates.[4] For percentage bias we only show the standard deviation.

We can use several criteria for judging the performance of the estimation methods. One is to look at the bias another the root mean squared error. Denoting the root mean squared error by RMSE and the standard deviation by std, we have the relation $RMSE^2 = std^2 + bias^2$. The

[4] Let $\overline{h}^{true}$ be the true mean of hours of work for a given tax system and $\overline{h}^{est}$ an estimate using the DC or BN method. Then the root mean squared error is defined as $\sqrt{\frac{1}{100}\sum_{k=1}^{100}\left(\overline{h}^{true} - \overline{h}_k^{est}\right)^2}$ and the standard deviation as $\sqrt{\frac{1}{100}\sum_{k=1}^{100}\left(\overline{\overline{h}}^{est} - \overline{h}_k^{est}\right)^2}$. The value for $\overline{h}^{true}$ is defined as the average hours of work for a sample consisting of 10,000,000 individuals.

RMSE can be high either because of a large variance in the estimates or because of a bias. Below we will comment on both the bias of the estimates and the RMSE.

Table 1 Translog utility function with parameter vector $\gamma\beta$ where γ is very large (equivalent to no GEV errors) and $\beta = (175 \quad 380 \quad -7 \quad -10 \quad 0)$, fixed preferences, no measurement errors

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2039.0	2037.5 (3.4) (3.0)	2038.2 (2.7) (2.6)	2042.9 (4.8) (2.9)	2043.1 (5.0) (2.9)
Percentage Bias	0	-0.07% (0.0015)	-0.04% (0.0013)	0.19% (0.0014)	0.20% (0.0014)
Hours tax system 5	1987.6	1986.9 (3.0) (2.9)	1987.4 (2.9) (2.9)	1982.1 (6.3) (3.0)	1983.5 (5.1) (3.0)
Percentage Bias	0	-0.04% (0.0014)	-0.01% (0.0015)	-0.28% (0.0015)	-0.21% (0.0015)
Change in hours Measure one	-51.4	-52.3 (3.2) (3.1)	-51.8 (3.2) (3.2)	-56.7 (6.1) (3.1)	-55.3 (4.9) (3.1)
Change in hours Measure two	-51.4	-50.7 (3.5) (3.4)	-50.8 (3.4) (3.3)	-60.8 (9.9) (3.0)	-59.6 (8.7) (3.0)
Percentage change measure one	-2.52%	-2.56% (0.0016) (0.0015)	-2.54% (0.0016) (0.0016)	-2.78% (0.0030) (0.0015)	-2.71% (0.0024) (0.0015)
Percentage change measure two	-2.52%	-2.49% (0.0017) (0.0017)	-2.49% (0.0016) (0.0016)	-2.98% (0.0048) (0.0015)	-2.92% (0.0042) (0.0015)

In table 1 we see that the DC predictions of average hours of work for both tax systems are very good, whether we add GEV errors in the prediction or not. Since the GEV errors are not important in the true data generating process it is reasonable that they are not important in the prediction. The predictions of the change in hours of work are also very good for the DC method. Since the likelihood function builds on the properties of the true data generating process these results are not surprising. The results confirm what we already know, maximum likelihood yields consistent estimates if the likelihood function is correctly specified. The performance of the BN method is also good, but not as good as the performance of the DC method. The predictions using the Lasso version are slightly better than those of the OLS version.

In table 2 we show results for a data generating process where the GEV errors play a role, although a minor one. The true hours of work are slightly lower than in table 1, where the β vector was scaled so that the GEV errors was of no importance. The DC predictions where GEV errors are added in the prediction are marginally better than when GEV errors are not added. Since the DC predictions are very accurate it does not matter whether the change in hours of work are calculated according to Measure one or Measure two. The DC method performs better than the BN method and the Lasso version is marginally better than the OLS version of the BN method. For the BN method, it is better to use the average hours of work for the sample for tax system 4 than the predicted value, when predicting the change in hours of work. The average hours of work for the sample for tax system 4 is an unbiased estimate of true hours, whereas it seems as if the BN-predictions are slightly biased, and in different directions for tax systems 4 and 5.

Table 2 Translog utility function plus GEV errors, fixed preferences, no measurement errors, $\beta = (175 \ \ 380 \ \ \text{-}7 \ \ \text{-}10 \ \ 0)$

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2037.9	2038.1 (3.0) (3.0)	2039.1 (2.9) (2.6)	2041.9 (5.0) (3.0)	2042.1 (5.2) (3.0)
Percentage Bias	0	0.01% (0.0015)	0.06% (0.0013)	0.19% (0.0015)	0.21% (0.0015)
Hours tax system 5	1986.9	1987.1 (3.0) (3.0)	1987.9 (3.2) (3.1)	1980.9 (6.7) (3.1)	1982.5 (5.4) (3.1)
Percentage Bias	0	0.01% (0.0015)	0.05% (0.0016)	-0.30% (0.0016)	-0.22% (0.0016)
Change in hours Measure one	-51.0	-50.9 (3.1) (3.1)	-50.1 (3.2) (3.1)	-56.9 (6.6) (3.0)	-55.3 (5.3) (3.1)
Change in hours Measure two	-51.0	-51.0 (3.4) (3.4)	-51.2 (3.2) (3.2)	-60.9 (10.3) (2.9)	-59.6 (9.1) (2.9)
Percentage change measure one	-2.50%	-2.50% (0.0015) (0.0015)	-2.46% (0.0016) (0.0015)	-2.79% (0.0032) (0.0015)	-2.71% (0.0026) (0.0015)
Percentage change measure two	-2.50%	-2.50% (0.0016) (0.0016)	-2.51% (0.0015) (0.0015)	-2.98% (0.0050) (0.0014)	-2.92% (0.0044) (0.0014)

The parameter vector used for the simulations in table 3 is $\beta =$ (8.75 19 -0.35 -0.5 0), i.e., the parameter vector used for the simulations in table 2 multiplied by 0.05, making the GEV errors more important for the distribution of utilities and hours of work. We see that making the GEV errors more important leads to a substantial decrease in the true average hours worked; about 45 hours for tax system 4 and 38 hours for tax system 5. The conclusion we should make is not that making the GEV errors more important decreases hours of work, but that the effect can be large.

Looking at the results in table 3, we see that for the DC method it now matters quite a lot whether GEV errors are added or not in the prediction. Not adding GEV errors in the prediction results in a large upward bias in hours of work for both tax system 4 and 5, while if GEV errors are added the average bias in the predictions of hours of work for tax systems 4 and 5 are small. That the predictions with the DC method adding GEV errors are good, is not surprising; the utility function is estimated with a correctly specified likelihood function and the prediction includes GEV errors just like in the true dgp. It is worth noting that the upward bias in the DC predictions of hours of work for tax systems 4 and 5 without GEV errors are of the same magnitude. This means that when calculating the change in hours when going from tax system 4 to 5 the bias cancels out to a large extent if the calculation of the change is done by using the predicted hours for tax system 4 (Measure 2). The BN estimators perform slightly better than in table 2; the Lasso version marginally better than the OLS version.

Table 3 Translog utility function plus GEV errors, fixed preferences, no measurement errors $\beta = (8.75\ \ 19\ \ -0.35\ \ -0.5\ 0)$

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	1994.5	1994.9 (6.3) (6.3)	2039.2) (44.9) (4.4)	1997.0 (5.9) (5.4)	1997.1 (6.0) (5.4)
Percentage Bias	0	0.02% (0.0032)	2.24% (0.0022)	0.12% (0.0027)	0.13% (0.0027)
Hours tax system 5	1950.1	1949.0 (6.2) (6.1)	1988.9 (38.1) (4.5)	1945.2 (6.8) (4.8)	1946.5 (5.9) (4.7)
Percentage Bias	0	-0.06% (0.0031)	1.94% (0.0023)	-0.25% (0.0024)	-0.18% (0.0024)
Change in hours Measure one	-44.4	-45.4 (7.3) (7.3)	-7.1 (37.7) (5.7)	-49.2 (6.1) (3.7)	-47.9 (5.1) (3.7)
Change in hours Measure two	-44.4	-45.9 (7.2) (7.1)	-51.3 (7.7) (3.4)	-51.8 (8.1) (3.4)	-50.6 (7.1) (3.4)
Percentage change measure one	-2.22%	-2.28% (0.0036) (0.0036)	-0.36% (0.0189) (0.0028)	-2.47% (0.0030) (0.0018)	-2.40% (0.0025) (0.0018)
Percentage change measure two	-2.22%	-2.30% (0.0036) (0.0035)	-2.52% (0.0033) (0.0016)	-2.59% (0.0040) (0.0017)	-2.53% (0.0035) (0.0017)

Summing up the results in tables 1-3 we see that, as the GEV errors become more important in the data generating process, there is a substantial effect on the true average hours of work, the bias of the BN estimates get slightly smaller, the bias of the DC2 estimator increases, the standard deviation and the RMSE increases for all the estimators. It is worth noting that the standard deviation increases more for DC1, the best performing DC estimator, than for the BN estimators. In fact, in table 3, the RMSE for the BN-Lasso is lower on all accounts than the RMSE for the DC estimators. This is a bit surprising, since the DC1 predictions build on estimates from a likelihood function that agrees with how data are generated and GEV errors are added in the prediction just as in the dgp.

The simulations in tables 1-3 have all been done with the same utility function as assumed when deriving the likelihood function and the predictions for DC1 are done exactly as assumed in the dgp assumed when deriving the likelihood function. In the next simulations we will use another utility function and generate data without GEV errors.[5] The utility function we will use is given by

$$u(c,h) = \exp\left\{-\left(1+\frac{\beta(c+\bar{s})}{b-h}\right)\right\}\left(\frac{h-b}{\beta}\right) \tag{4}$$

where *c* is consumption, *h* hours of work, $\bar{s} = (s/\beta) - (\alpha/\beta^2)$. For parameter values such that preferences are convex, maximization of this utility function subject to a linear budget constraint with slope *w* and intercept *y*, yields the labor supply function:

$$h = s + \alpha w + \beta y \tag{5}$$

This is a functional form for the labor supply function that has been used extensively in the labor supply literature. When specifying the parameter values for this labor supply (utility) function we do this for hours of work measured in thousands and consumption in thousands of kronor. The choice of parameter values is inspired by empirical results in Blomquist (1983) and Blomquist and Hansson-Brusewitz (1990).

In our first simulation with this utility (labor supply) function we let the preferences be fixed ($s = 1.25$, $\alpha = 0.014$ and $\beta = -0.0004$). This means that the only difference between the results presented in table 4 and those presented in table 1 is that, as seen from the perspective of DC estimation, the function used to derive the log likelihood function is misspecified; this is the first change in the dgp that makes the likelihood function misspecified. Results are shown in table 4. Looking first at the performance of the two DC estimators we see that the bias in the predictions of hours of work for both tax systems shoot up, as do the Root mean squared errors. The predictions of the change in hours of work is also worse than when the dgp generating data was the same as the one assumed when constructing the likelihood function. However, since

[5] I regard the assumption of GEV errors in the dgp as a technical assumption made to motivate the form of the likelihood function, not because there are any data that suggest that this really is the data generating process. In the following I will therefore use data generating processes that do not include GEV errors. An alternative interpretation is that the utility values given by the quasi-concave utility function are very large in comparison to the GEV errors, so that the GEV errors have no influence on how hours of work are chosen.

the bias in the predictions for tax system 4 and tax system 5 is in the same direction, the bias partly cancels when using Measure two for the prediction of the change in hours of work.

The BN method does not assume a particular form for the utility function, it only assumes preferences are convex. If we should talk about misspecification when discussing the BN method it would be that for all results in tables 1-4 preferences have been fixed, whereas the BN method is designed to handle the case where preferences are heterogenous across individuals. Comparing the performance of the BN estimators between tables 1 and 4 we see that the bias in the prediction of hours of work for tax system 4 goes down, but that it goes up in the predictions for tax system 5. The Root mean squared errors for the change in hours of work are higher when we use utility function (4) to generate data than when we use the translog utility function.

Table 4 Discrete dgp, linear supply function, fixed preferences, no measurement errors

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2005.7	1989.1 (16.8) (2.8)	1991.5 (14.4) (2.6)	2003.9 (3.1) (2.6)	2003.7 (3.3) (2.6)
Percentage error	0%	-0.83% (0.0014)	-0.71% (0.0013)	-0.09% (0.0013)	-0.10% (0.0013)
Hours tax system 5	1957.0	1946.3 (11.0) (2.6)	1949.1 (8.3) (2.6)	1943.5 (13.7) (2.3)	1943.7 (13.4) (2.3)
Percentage error	0%	-0.54% (0.0013)	-0.40% (0.0013)	-0.69% (0.0012)	-0.68% (0.0012)
Change in hours Measure one	-48.7	-59.2 (10.6) (1.4)	-56.4 (7.8) (1.2)	-62.0 (13.3) (0.7)	-61.7 (13.0) (0.7)
Change in hours Measure two	-48.7	-42.8 (6.1) (1.4)	-42.4 (6.4) (1.1)	-60.4 (11.7) (0.7)	-59.9 (11.2) (0.6)
Percentage change measure one	-2.43%	-2.95% (0.0053) (0.0007)	-2.81% (0.0039) (0.0006)	-3.09% (0.0066) (0.0003)	-3.08% (0.0065) (0.0003)
Percentage change measure two	-2.43%	-2.15% (0.0028) (0.0007)	-2.13% (0.0031) (0.0006)	-3.01% (0.0059) (0.0003)	-2.99% (0.0056) (0.0003)

In the next set of simulations, we will take explicit account of preference heterogeneity and let s, β and α vary across individuals. Individuals choose hours of work according to the rule used in the discrete choice labor supply literature, as described by eq. (2), but without adding GEV errors. Remembering that, when we state values of the preference parameters we always do that with the understanding that labor supply is measured in thousands of hours and consumption in hundred thousands of kronor, we use the following parameter values; the coefficients s_i are drawings from a normal distribution with mean 1.25 and standard deviation 0.12, the wage rate coefficients α_i are drawings from a truncated normal distribution; the untruncated distribution having mean 0.014 and standard deviation 0.0015, the lower truncation being 0.008 and the upper truncation being 0.020. The mean of the truncated distribution is 0.014 and the standard deviation 0.0015. The nonlabor income coefficients β_i are drawings from a truncated normal distribution; the untruncated distribution having mean 0.2 and standard deviation 0.009, the lower truncation being -0.002 and the upper zero. The mean of the truncated distribution is -0.00039 and the standard deviation 0.00037. [6]

Table 5 Discrete dgp, linear supply function, random preferences; no measurement errors.

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1991.8 (17.8) (3.9)	2007.7 (3.9) (3.7)	2009.7 (3.6) (3.5)	2009.9 (3.6) (3.5)
Percentage Bias	0	-0.80% (0.0020)	-0.07% (0.0018)	0.03% (0.0018)	0.04% (0.0018)
Hours tax system 5	1952.1	1947.5 (5.9) (3.7)	1965.1 (13.4) (3.5)	1947.0 (6.1) (3.3)	1947.6 (5.6) (3.3)
Percentage Bias	0	-0.24% (0.0019)	0.67% (0.0018)	-0.26% (0.0017)	-0.23% (0.0017)
Change in hours Measure one	-57.0	-61.3 (4.7) (2.1)	-43.7 (13.5) (1.4)	-61.8 (4.8) (0.8)	-61.2 (4.2) (0.8)
Change in hours Measure two	-57.0	-44.3 (13.0) (2.7)	-42.6 (14.5) (1.2)	-62.7 (5.7) (0.7)	-62.3 (5.3) (0.7)
Percentage change measure one	-2.84%	-3.05% (0.0024) (0.0010)	-2.17% (0.0067) (0.0007)	-3.07% (0.0024) (0.0004)	-3.05% (0.0021) (0.0004)
Percentage change measure two	-2.84%	-2.22% (0.0063) (0.0013)	-2.12% (0.0072) (0.0006)	-3.12% (0.0028) (0.0004)	-3.10% (0.0026) (0.0003)

[6] There have been attempts to modify the DC likelihood function so it takes account of heterogenous preferences in the $u(c,h)$ in the function given by eq. (5), but without much success. See more on this in Appendix A.

It is interesting to compare the results in tables 4 and 5. Starting with the DC estimators we see that the within sample prediction, tax system 4, is about the same for DC1 in both tables, whereas the ability to predict out of sample, tax system 5, is better when there is preference heterogeneity. For DC2 the performance to predict for tax system 4 is much better when we have heterogenous preferences, but the out of sample prediction is worse and goes from a negative bias to a positive bias. In terms of predicting the change in hours of work DC1 is best whether the preferences are fixed or random. However, when preferences are fixed it is best to use Measure 2, whereas if preferences are random it is better to use Measure 1. The bias of the BN estimators is less in table 5 and as a consequence the RMSE are lower when the preferences are random, with the exception of the BN-OLS prediction for tax system 4. BN-Lasso using Measure one gives the best prediction of the percentage change in hours of work and the RMSE is slightly lower than the RMSE for the best DC estimator. We note that the BN-Lasso performs better than the DC estimators, both in terms of bias and RMSE.

The BN-method is designed to handle both heterogeneous preferences and measurement errors. In table 6 we show results when we have both random preferences and measurement errors. The measurement errors are drawings from a normal distribution with mean zero and standard deviation 200 hours. In tables 7-10 we successively increase the standard deviation in steps of 200, so that in table 10 the standard deviation is 1000.

Table 6 Discrete dgp, linear supply function, random preferences, measurement error std 200

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1989.8 (20.0) (4.7)	2034.4 (25.5) (3.7)	2009.4 (4.2) (4.2)	2009.6 (4.2) (4.2)
Percentage bias	0	-0.97% (0.0024)	1.25% (0.0019)	0.01% (0.0021)	0.02% (0.0021)
Hours tax system 5	1952.1	1945.5 (7.8) (4.1)	1992.5 (40.5) (3.8)	1946.7 (6.7) (3.9)	1947.3 (6.2) (3.8)
Percentage bias	0	-0.34% (0.0021)	2.07% (0.0019)	-0.28% (0.0020)	-0.25% (0.0020)
Change in hours Measure one	-57.0	-62.9 (6.9) (3.5)	-16.0 (41.1) (2.3)	-61.8 (5.0) (1.4)	-61.2 (4.4) (1.4)
Change in hours Measure two	-57.0	-44.2 (13.4) (3.9)	-41.9 (15.2) (1.2)	-62.7 (5.8) (1.1)	-62.3 (5.4) (1.1)
Percentage change measure one	-2.84%	-3.13% (0.0034) (0.0017)	-0.80% (0.0205) (0.0011)	-3.08% (0.0025) (0.0007)	-3.05% (0.0022) (0.0007)
Percentage change measure two	-2.84%	-2.22% (0.0065) (0.0019)	-2.06% (0.0078) (0.0006)	-3.12% (0.0029) (0.0005)	-3.10% (0.0027) (0.0005)

Before commenting on how results change as we increase the standard deviation of the measurement error it can be worthwhile to give some detailed comments on the results in table 6. If we compare the performance of DC1 and DC2 we see that DC1 is a clear winner. Whatever prediction we look at, DC1 has a smaller bias and a smaller Root mean squared error than DC2. If we compare BN-OLS and BN-Lasso we see that their performances are very similar, but that the BN-Lasso is marginally better. The prediction of the change in hours of work is best if we use Measure 1 for the prediction. If we compare BN-Lasso with DC1, the best DC estimator, we see that BN-Lasso has both a lower bias and a lower Root mean squared error whatever prediction we look at.

It can also be of interest to comment on how results change as we increase the standard deviation of the measurement error from zero (table 5) up to 1000 hours (table 10). In table 5, where we show results for simulations without measurement errors, we see that DC1 has a downward percentage bias of -0.80% for the prediction of hours of work for tax system 4 and -0.24% for tax system 5, the RMSE is 17.8 for the prediction for tax system 4 and 5.9 for tax system 5. When measurement errors with standard deviation 200 are added to hours of work these downward biases increases to -0.97% and -0.34% and the RMSE to 20.0 and 7.8. As the standard deviation of the errors increases the downward bias increases successively to reach a downward bias of -5.7% for tax system 4 and -4.65% for tax system 5 and with RMSE 115.8 and 91.3 respectively, when the standard deviation for the optimization/measurement error is 1000. Since the bias in the predictions for tax systems 4 and 5 are in the same direction, the bias cancels out to some extent when we calculate the change in hours of work using Measure 2. Still, the prediction of the change in hours of work is worse than the predictions obtained by the BN-method. We see in table 5 that there is a very small downward bias of -0.07% in the prediction for tax system 4 and an upward bias of 0.67% for tax system 5 when we use DC2 for the prediction and that the RMSE is 3.9 for the prediction for tax system 4 and 13.4 for tax system 5. Adding measurement errors to hours of work induces an upward bias in the estimates using DC2. When the standard deviation of the measurement errors is 200 the upward bias is 1.25% for the prediction for tax system 4 and 2.07% for tax system 5 and the RMSE is 25.5 and 40.5 respectively. These biases increase successively as the standard deviation increases and reaches 44.38% for tax system 4 and 46.75% for tax system 5 when the standard deviation is 1000 and RMSE 891.8 and 912.8 respectively. The bias in the DC2 predictions of hours of work for tax systems 4 and 5 are in the same direction, which means that when we calculate the change in hours of work according to Measure 2, the bias cancels out to some extent. Still, the predictions of the change in hours of work is not good.

When we use the BN-method, the percentage bias of the in-sample prediction for tax system 4 stays pretty much the same as we increase the standard deviation of the measurement error and is around 0 to 0.05%. The bias for the out of sample prediction for tax system 5 stays also roughly the same and is around 0.25% to 0.30%. However, as we increase the dispersion of the measurement errors, the Root Mean Squared Error increases, this increase is because the standard deviation of the estimates increases. In general, the performance of the BN-Lasso is marginally better than the BN-OLS. The prediction of the change in hours work is best for the BN-Lasso using Measure 1. The true change in hours of work is 2.84% and the BN-Lasso, Measure 1, predicts this change to be 3.05%, irrespective of the standard deviation of the measurement error. However, the RMSE of the prediction increases as the standard deviation of the measurement error increases.

The dgp:s generating the data for tables 5-10 are all the same, except that the standard deviation of the optimization/measurement error is successively increased, from zero in table 5 to 1000 in table 10. It seems safe to say that zero is too low and 1000 too high. Since, the properties of the estimators change as the standard deviation increases, and especially so for the DC estimators, it would be of great value if we could give an interval that we believe contains the standard deviation of the optimization/measurement error in real data. Unfortunately, I have not come across information that allows me to give such an interval.

Maybe one can get a rough idea of the variance of the optimization errors in hours of work by looking at the variance of the optimization errors for taxable income. Lack of spikes at kink points of the budget constraint and the lack of holes in the distribution of taxable income at notches could maybe be used to get an estimate of the variance of the optimization errors for taxable income. However, I do not know if there have been attempts to obtain estimates of the variance of optimization errors in taxable income by looking at the distribution of taxable income at kinks or notches.

Studies of labor supply usually uses survey data. Often the survey questions ask about hours of work the previous year. There can be measurement errors for many reasons; respondents might misinterpret the questions, there might be problems of recall etc. There have been attempts to validate these kinds of data by comparing them with company data, see e.g. Duncan and Hill (1985). Duncan and Hill regard the company data as true data, which is problematic since the company data are constructed to measure something different than what we want to use in labor supply studies, e.g. the company data has no information on hours of work in places outside the company in question. Moreover, according to Duncan and Hill (1985), overtime hours are sometimes poorly recorded in the company data. There are models estimated by maximum likelihood methods that do allow for measurement errors in hours of work and generate estimates of the variance of the measurement error, see e.g. Heim (2009). It is hard to draw any strong conclusions about the variance of the measurement error from these studies.

Table 7 Discrete dgp, linear supply function, random preferences, optimization/measurement error std 400

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1981.3 (28.5) (6.1)	2090.1 (81.1) (4.2)	2009.1 (5.4) (5.4)	2009.2 (5.4) (5.4)
Percentage Bias	0	-1.39% (0.0030)	4.03% (0.0021)	≈ 0% (0.0027)	≈ 0% (0.0027)
Hours tax system 5	1952.1	1937.2 (16.0) (5.9)	2046.5 (94.5) (4.1)	1946.4 (7.6) (5.0)	1947.0 (7.1) (4.9)
Percentage Bias	0	-0.76% (0.0030)	4.84% (0.0021)	-0.29% (0.0026)	-0.26% (0.0025)
Change in hours Measure one	-57.0	-71.0 (15.2) (6.1)	38.4 (95.5) (4.1)	-61.8 (5.3) (2.3)	-61.2 (4.8) (2.3)
Change in hours Measure two	-57.0	-44.1 (14.3) (6.1)	-43.6 (13.5) (1.4)	-62.7 (6.0) (1.8)	-62.3 (5.5) (1.8)
Percentage change measure one	-2.84%	-3.53% (0.0075) (0.0030)	1.91% (0.0475) (0.0021)	-3.08% (0.0026) (0.0011)	-3.05% (0.0024) (0.0011)
Percentage change measure two	-2.84%	-2.33% (0.0068) (0.0031)	-2.09% (0.0076) (0.0007)	-3.12% (0.0030) (0.0009)	-3.10% (0.0027) (0.0009)

Table 8 Discrete dgp, linear supply function, random preferences, measurement error std 600

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1960.3 (49.4) (7.2)	2163.0 (153.9) (5.4)	2008.8 (6.8) (6.8)	2008.9 (6.8) (6.8)
Percentage Bias	0	-2.43% (0.0036)	7.66% (0.0027)	-0.02% (0.0034)	-0.02% (0.0034)
Hours tax system 5	1952.1	1981.9 (34.1) (7.5)	2118.0 (169.5) (5.5)	1946.0 (8.8) (6.3)	1946.6 (8.2) (6.2)
Percentage Bias	0	-1.7% (0.0039)	8.50% (0.0028)	-0.31% (0.0032)	-0.28% (0.0032)
Change in hours Measure one	-57.0	-89.0 (33.1) (8.7)	110.1 (167.2) (5.9)	-61.8 (5.8) (3.3)	-61.2 (5.3) (3.3)
Change in hours Measure two	-57.0	-41.4 (18.1) (9.2)	-45.0 (12.1) (1.7)	-62.8 (6.3) (2.6)	-62.3 (5.8) (2.5)
Percentage change measure one	-2.84%	-4.43% (0.0165) (0.0043)	5.48% (0.0833) (0.0031)	-3.08% (0.0029) (0.0016)	-3.05% (0.0026) (0.0016)
Percentage change measure two	-2.84%	-2.11% (0.0086) (0.0047)	-2.08% (0.0076) (0.0008)	-3.12% (0.0031) (0.0012)	-3.10% (0.0029) (0.0012)

Table 9 Discrete dgp, linear supply function, random preferences, measurement error std 800

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1929.2 (80.4) (8.6)	2361.6 (352.5) (9.9)	2008.5 (8.4) (8.4)	2008.6 (8.4) (8.4)
Percentage error	0	-3.98% (0.0043)	17.54% (0.0049)	-0.03% (0.0042)	-0.03% (0.0042)
Hours tax system 5	1952.1	1891.8 (61.1) (9.7)	2307.6 (355.6) (9.7)	1945.7 (10.0) (7.7)	1946.3 (9.6) (7.6)
Percentage error	0	-3.09% (0.0050)	18.21% (0.0050)	-0.33% (0.0639)	-0.30% (0.0039)
Change in hours Measure one	-57.0	-115.8 (59.7) (10.8)	300.0 (357.2) (9.8)	-61.8 (6.4) (4.3)	-61.3 (6.0) (4.3)
Change in hours Measure two	-57.0	-37.4 (22.1) (10.2)	-54.0 (4.3) (3.0)	-62.8 (6.6) (3.3)	-62.3 (6.2) (3.3)
Percentage change measure one	-2.84%	-5.76% (0.0297) (0.0053)	14.95% (0.1779) (0.0052)	-3.08% (0.0032) (0.0021)	-3.05% (0.0030) (0.0021)
Percentage change measure two	-2.84%	-1.94% (0.0104) (0.0052)	-2.29% (0.0057) (0.0016)	-3.13% (0.0033) (0.0016)	-3.10% (0.0031) (0.0016)

Table 10 Discrete dgp, linear supply function, random preferences, measurement error, std 1000

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1893.7 (115.8) (9.2)	2900.8 (891.8) (17.6)	2008.2 (10.0) (10.0)	2008.3 (10.0) (10.0)
Percentage Bias	0	-5.75% (0.0046)	44.38% (0.0088)	0.05% (0.0050)	0.04% (0.0050)
Hours tax system 5	1952.1	1861.4 (91.3) (10.6)	2864.7 (912.8) (20.0)	1945.4 (11.3) (9.2)	1946.0 (10.9) (9.1)
Percentage Bias	0	-4.65% (0.0054)	46.75% (0.0102)	-0.34% (0.0047)	-0.32% (0.0047)
Change in hours Measure one	-57.0	-145.9 (89.7) (12.4)	857.5 (914.7) (20.2)	-61.8 (7.2) (5.3)	-61.3 (6.8) (5.3)
Change in hours Measure two	-57.0	-32.3 (27.4) (11.8)	-36.0 (21.4) (4.3)	-62.8 (7.1) (4.1)	-62.4 (6.7) (4.0)
Percentage change measure one	-2.84%	-7.26% (0.0447) (0.0060)	42.72% (0.4557) (0.0110)	-3.08% (0.0035) (0.0026)	-3.05% (0.0033) (0.0025)
Percentage change measure two	-2.84%	-1.70% (0.0129) (0.0062)	-1.24% (0.0160) (0.0015)	-3.13% (0.0035) (0.0020)	-3.11% (0.0033) (0.0020)

Quantity constraints

Many individuals probably face quantity constraints. An important question is how to model this. In the BN method this is captured by assuming there is some desired hours of work, h^*. Because of quantity constraints an individual might not be able to work these number of hours, but has to work some other hours. This can be represented by a term η so that actual hours become $\tilde{h} = h^* + \eta$. Here, η represents the difference between desired hours and the actual hours, an optimization error. Of course, there can also be measurement errors. Using ε to denote a measurement error, observed hours can be written as $h = \tilde{h} + \varepsilon = h^* + \eta + \varepsilon$. In the discrete choice literature, one represents the quantity constraints by a fixed set of discrete points of hours of work an individual can chose from. It is assumed this set is the same for everyone. We can have two different interpretations of this. One is that some scholars actually believe there are fixed constraints like this, the same for everyone. Another interpretation is that it simply is a convenient way to represent nonlinear budget sets. Irrespective of interpretation, the likelihood function is constructed under the assumption of a given set of discrete points, the same for everyone. A way to move closer to reality in the data generating process is to let individuals face different sets of fixed points they can chose from. In the next simulation we will keep the assumption that each individual face 11 points to choose from. Zero and 3000 hours will be in the choice set for all individuals, the other 9 points will be random drawings from a uniform distribution defined on the interval [150, 2850]. In a second simulation we will let the quantity constraints be represented by 29 discrete points. In table 11 we show the results obtained when individuals choose from 11 points, but where the choice sets differ between individuals. In table 12 we give the results where individuals can choose between 29 points, but where the choice set differs across individuals.

Comparing the results in table 11 with those in table 5, we see that the true hours of work are somewhat higher in table 11. The BN method can still predict average hours of work for tax systems 4 and 5 quite precisely and the performance in terms of RMSE is about the same as when all individuals face the same 11 points to choose from. The performance of DC1 is similar to the performance when individuals choose from the same 11 points. However, the performance of DC2 is much worse. Looking at table 12 we see that true hours are pretty close to those in table 5 and the performance of the BN-method is very similar to that shown in table 5. Likewise, the predictions of DC1 are quite close to those shown in table 5. The bias of DC2 is not as bad as that shown in table 11, but still worse than when all individuals choose from the same 11 points.

Table 11 Discrete dgp, linear supply function, random preferences, no measurement error, 2+9 random points

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2014.5	1997.6 (17.5) (4.3)	2038.4 (24.1) (3.5)	2016.2 (4.4) (4.1)	2016.4 (4.5) (4.1)
Percentage Bias	0%	-0.84% (0.0021)	1.19% (0.0017)	0.09% (0.0020)	0.10% (0.0020)
Hours tax system 5	1956.1	1952.1 (6.9) (4.4)	1996.0 (40.1) (3.7)	1952.3 (5.2) (3.7	1953.0 (4.8) (3.7)
Percentage Bias	0%	-0.20% (0.0023)	2.04% (0.0019)	-0.19% (0.0019)	-0.16% (0.0019)
Change in hours Measure one	-58.4	-63.1 (5.7) (3.4)	-19.1 (39.4) (2.2)	-62.8 (4.6) (1.4)	-62.2 (4.0) (1.4)
Change in hours Measure two	-58.4	-45.5 (13.5) (3.9)	-42.4 (16.1) (1.3)	-63.9 (5.6) (1.2	-63.5 (5.2) (1.2)
Percentage change measure one	-2.90%	-3.13% (0.0028) (0.0017)	-0.95% (0.0196) (0.0011)	-3.12% (0.0027) (0.0007)	-3.08% (0.0020) (0.0007)
Percentage change measure two	-2.90%	-2.28% (0.0065) (0.0019)	-2.08% (0.0083) (0.0007)	-3.17% (0.0027) (0.0006)	-3.15% (0.0025) (0.0006)

Table 12 Discrete dgp, linear supply function, random preferences, no measurement error, 2+27 random points

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.9	1992.4 (17.0) (3.9)	2011.8 (4.6) (3.6)	2010.3 (3.9) (3.6)	2010.6 (4.0) (3.6)
Percentage Bias	0%	-0.83% (0.0019)	0.14% (0.0018)	0.07% (0.0018)	0.08% (0.0018)
Hours tax system 5	1951.6	1947.8 (5.7) (4.2)	1969.1 (17.9) (3.8)	1947.1 (5.7) (3.5)	1947.8 (5.2) (3.5)
Percentage Bias	0%	-0.20% (0.0022)	0.89% (0.0020)	-0.23% (0.0018)	-0.20% (0.0018)
Change in hours Measure one	-57.3	-61.5 (4.8) (2.3)	-40.2 (17.2) (1.4)	-62.2 (5.0) (0.9)	-61.5 (4.3) (0.9)
Change in hours Measure two	-57.3	-44.6 (13.1) (3.0)	-42.7 (14.6) (1.2)	-63.2 (6.0) (0.8)	-62.8 (5.5) (0.8)
Percentage change measure one	-2.85%	-3.06% (0.0024) (0.0011)	-2.00% (0.0085) (0.0007)	-3.10% (0.0025) (0.0004)	-3.06% (0.0021) (0.0004)
Percentage change measure two	-2.85%	-2.24% (0.0063) (0.0015)	-2.12% (0.0073) (0.0006)	-3.15% (0.0029) (0.0004)	-3.13% (0.0027) (0.0004)

I have also done simulations with 100 points, zero hours, 3000 hours and 98 randomly (uniform distribution) distributed points. In practice this dgp should be very close to a dgp where individuals can choose hours of work continuously. In fact, the results are very close to those obtained from a continuous dgp with random preferences and no measurement error. Since the results are in fact almost identical to those in table 16, I do not present them in a separate table.

Nonconvex budget constraints

It is sometimes said that one of the advantages of the DC method is that it can easily handle nonconvex budget constraints. To see to what extent this is true I have performed simulations where the budget constraints have a nonconvex portion. To construct the nonconvex budget constraints, I modify the budget constraints used for the simulations above. I describe the modifications in the taxable income consumption space. First, I add 30,000 crowns as a lump-sum income to all budget constraints. We keep the kink points as for the previous simulations, but now the first segments have low slopes varying between 0.5 and 0.6 for the various tax systems, implying marginal tax rates between 40 and 50%. The slopes of the second segments vary between 0.75 and 0.9, implying marginal tax rates between 10 and 25%. This generates budget constraints where the two first segments create a nonconvex part of the budget constraint. The slopes for the third and fourth segments are as in the previous simulations and hence create a convex part of the budget constraints. To exemplify, for tax system 4 the vector of slopes was $s4 = [1.0 \;\; 0.8 \;\; 0.7 \;\; 0.6]$ but is now $s4 = [0.6 \;\; 0.8 \;\; 0.7 \;\; 0.6]$, and for tax system 5 the vector of slopes was $s5 = [1.0 \;\; 0.75 \;\; 0.65 \;\; 0.55]$ but is now $s5 = [0.55 \;\; 0.8 \;\; 0.65 \;\; 0.55]$. The kink points in terms of thousand of crowns are unchanged at $k4 = k5 = [0 \;\; 80 \;\; 160 \;\; 250]$. This choice of slopes generates budget constraints that from the kink point 160,000 is the same as in the earlier simulations. The difference in average hours for tax system 5 as compared with the earlier simulations is due to the nonconvex part. The share of individuals that have an optimum on the nonconvex part of the budget constraint varies between tax systems, but is in general around 45%.

The estimating functions used by the BN method builds on the assumption that the budget constraints are convex, but in the simulations presented below the budget constraints are nonconvex. This implies that the functional forms used for the BN estimations are misspecified, unless we modify the functional forms used for the estimation. More on this below. There is nothing in the DC setup that prescribes that the budget constraints must be convex. Hence, the degree of misspecification for the DC method is qualitatively the same as for the simulations above. The simulations are done with the utility function given by equation (4) and with heterogenous preferences.

If the budget constraint is nonconvex, expression (B4) in Appendix B is no longer valid. However, as shown in Blomquist and Newey (2002) and Blomquist et al. (2023), if the nonconvex section of the budget constraint consists of two segments with slopes $s_{j+1} > s_j$ and with virtual incomes $y_{j+1} < y_j$, then expression (B4) can be extended with a term $\mu(s_j, s_{j+1}, y_j, y_{j+1})$ which takes care of the nonconvexity.[7] The $\mu(\;)$ term represents the deviation of the mean from what it would be if the budget constraint was convex. In the present simulations it is segments 1 and 2 that together constitute the nonconvex part of the budget constraint. In the estimations and predictions, I approximate $\mu(s_1, s_2, y_1, y_2)$ by a function with linear and quadratic terms of the arguments in $\mu(\;)$; no cross terms are included.

In table 13 we show results when there are no optimization/measurement errors and in table 14 when there are additive measurement errors with a standard deviation of 400.

[7] See Blomquist et al. (2023) p. 13 for more details.

Table 13 Discrete dgp, linear supply function, random preferences, non-convex budget constraints, no measurement errors

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2004.6	1979.4 (25.4) (3.4)	1995.5 (9.6) (3.0)	2005.5 (3.0) (2.9)	2003.9 (3.0) (2.9)
Percentage Bias	0	-1.26% (0.0017)	-0.45% (0.0015)	0.04% (0.0014)	-0.04% (0.0014)
Hours tax system 5	1956.0	1948.3 (8.2) (2.9)	1963.8 (8.3) (2.9)	1951.2 (5.8) (3.5)	1953.8 (3.5) (2.8)
Percentage Bias	0	-0.39% (0.0015)	0.40% (0.0015)	-0.25% (0.0017)	-0.11% (0.0014)
Change in hours Measure one	-48.7	-56.4 (8.1) (2.2)	-41.0 (7.8) (1.6)	-53.6 (5.6) (2.5)	-51.0 (2.9) (1.7)
Change in hours Measure two	-48.7	-31.1 (17.8) (2.6)	-31.7 (17.0) (1.3)	-54.4 (6.3) (2.7)	-50.1 (2.2) (1.7)
Percentage change measure one	-2.43%	-2.82% (0.0040) (0.0011)	-2.04% (0.0039) (0.0008)	-2.67% (0.0028) (0.0013)	-2.54% (0.0014) (0.0009)
Percentage change measure two	-2.43%	-1.57% (0.0087) (0.0013)	-1.59% (0.0084) (0.0006)	-2.71% (0.0031) (0.0013)	-2.50% (0.0011) (0.0009)

We can compare the results in Table 13 with those in Table 5. The differences are due to the fact that the results in Table 13 is for nonconvex budget constraints and those in Table 5 for convex budget constraints. Looking at the results for the DC estimators we see that the RMSE for DC1 is in general higher in table 13 than in Table 5. For DC2 it is the other way around, the RMSE are lower in Table 13 than in Table 5. The RMSE for the BN-OLS estimator is about the same in tables 5 and 13. The RMSE for the BN-Lasso estimator is lower in table 13 than in table 5 and performs better than the OLS version. Both BN-methods perform better than the DC estimators.

Table 14 shows results where an additive optimization/measurement error with standard deviation 400 has been added to individuals' desired hours. The results in Table 14 can be compared with those in Table 7. We see that the RMSE for the DC estimators are higher than in table 7 and that DC1 performs better than DC2. The RMSE for both BN estimators are higher in table 14 than in table 7. This is because the standard deviations are higher in table 14. Both

BN estimators outperform the DC estimators. From these simulations we conclude that the DC estimators perform better when budget constraints are convex. The simulations indicate that the BN estimators can handle nonconvex budget constraints better than the DC estimators.

Table 14 Discrete dgp, linear supply function, random preferences, non-convex budget constraints, measurement errors std=400

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2004.6	1965.7 (39.3) (5.7)	2094.1 (89.6) (4.1)	2005.5 (4.9) (4.9)	2004.1 (4.7) (4.7)
Percentage Bias	0	-1.94% (0.0028)	4.46% (0.0020)	0.04% (0.0024)	-0.03% (0.0023)
Hours tax system 5	1956.0	1938.2 (18.8) (6.0)	2063.8 (107.9) (4.0)	1950.5 (8.1) (6.0)	1953.8 (6.2) (5.9)
Percentage Bias	0	-0.91% (0.0031)	5.51% (0.0020)	-0.28% (0.0031)	-0.11% (0.0030)
Change in hours Measure one	-48.7	-66.7 (18.9) (5.7)	+58.9 (107.7) (3.7)	-54.4 (7.8) (5.4)	-51.1 (5.8) (5.2)
Change in hours Measure two	-48.7	-27.5 (22.0) (6.2)	-30.3 (18.4) (0.9)	-55.0 (8.3) (5.4)	-50.3 (5.3) (5.1)
Percentage change measure one	-2.43%	-3.33% (0.0094) (0.0028)	+2.94% (0.0537) (0.0019)	-2.71% (3.9) (2.7)	-2.55% (0.0029) (0.0026)
Percentage change measure two	-2.43%	-1.40% (0.0107) ((0.0031)	-1.45% (0.0098) (0.0004)	-2.74% (0.0041) (0.0027)	-2.51% (0.0026) (0.0025)

Sample size

All the simulations so far have been with a sample size of 40,000. It can be of interest to see how results change if we use a much smaller sample size. In table 15 I show results when we have data from 500 individuals for each period used, i.e. the sample size used for the estimation is 2,000.

Table 15 Discrete dgp, linear supply function, random preferences, measurement error std 400, sample size 2000

	True	DC1 with GEV errors	DC2 no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1978.6 (40.8) (27.1)	2088.1 (80.9) (18.0)	2007.4 (22.0) (22.0)	2007.4 (22.0) (22.0)
Percentage error	0	-1.52% (0.0135)	3.93% (0.0090)	-0.09% (0.0110)	-0.09% (0.0110)
Hours tax system 5	1952.1	1935.2 (33.0) (28.5)	2044.9 (94.5) (18.2)	1945.4 (21.5) (20.6)	1945.6 (21.3) (20.3)
Percentage error	0	-0.87% (0.0146)	4.75% (0.0093)	-0.35% (0.0105)	-0.33% (0.0104)
Change in hours Measure one	-57.0	-71.5 (30.5) (27.0)	38.1 (96.6) (17.0)	-61.4 (12.3) (11.5)	-61.2 (12.0) (11.4)
Change in hours Measure two	-57.0	-43.4 (32.6) (29.8)	-43.2 (15.0) (6.0)	-62.0 (10.0) (8.8)	-61.8 (9.9) (8.7)
Percentage change measure one	-2.84%	-3.56% (0.0151) (0.0133)	1.90% (0.0482) (0.0086)	-3.06% (0.0060) (0.0056)	-3.05% (0.0059) (0.0055)
Percentage change measure two	-2.84%	-2.18% (0.0162) (0.0149)	-2.07% (0.0082) (0.0029)	-3.09% (0.0049) (0.0043)	-3.08% (0.0048) (0.0042)

The standard errors, and hence the RMSE, increase for all estimators. The negative bias of DC1 in the prediction of hours of work for tax systems 4 and 5 increases somewhat as does the positive bias of DC2. For the BN-estimators there is now a small negative bias in the prediction of hours of work for tax system 4 and the negative bias in the prediction for tax system 5 increases. The prediction for hours of change are almost unchanged for all the estimators; the most notable change, as compared with the results for sample size 40,000, is the increase in the standard errors and root mean squared errors.

4. Simulations where individuals choose from a continuous budget constraint

In this section I present simulations where individuals can choose any point on a piece-wise linear continuous budget constraint. In table 16 I present results where data are generated with random preferences but no measurement errors. The data generating process is the same as the one used to generate the simulations shown in table 5, except for the fact that in the simulations shown in table 5 individuals could only choose from 11 points, in the present dgp they can choose any point on the budget constraint. We see that true hours are 0.9 hours lower for tax system for 4 and 1.1 hours lower for tax system 5 when individuals can choose hours of work freely on the budget constraint. The performance of the estimators as shown in table 13 is very similar to the performance shown in table 5. In table 17 we show results when normally distributed optimization/measurement errors with a standard deviation 200 are added to the desired hours of work. In tables 17 up to table 21 we show results where we successively increase the standard deviation of optimization/measurement errors so that in table 21 we show results where the standard deviation is 1000. Comparing the results in tables 16-21 with the corresponding tables where data has been generated with the discrete data generating process we see that results are very similar. If anything, the RMSE is slightly lower when the continuous dgp has been used to generate the data. We conclude that the relative performance of the DC and BN estimators does not depend on whether we use the discrete or the continuous dgp to generate the data.

Table 16 Continuous dgp, linear supply function, random preferences, no measurement error,

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1991.5 (17.3) (3.7)	2007.7 (3.4) (3.3)	2009.4 (3.3) (3.2)	2009.7 (3.4) (3.2)
Percentage Bias	0	-0.84% (0.0018)	-0.03% (0.0017)	0.05% (0.0016)	0.07% (0.0016)
Hours tax system 5	1951.0	1947.5 (4.9) (3.5)	1964.9 (14.3) (3.2)	1946.3 (5.5) (3.0)	1947.0 (4.9) (2.9)
Percentage Bias	0	-0.18% (0.0018)	0.71% (0.0017)	-0.24% (0.0015)	-0.20% (0.0015)
Change in hours Measure one	57.4	-60.7 (3.9) (2.1)	-43.4 (14.1) (1.5)	-62.0 (4.7) (0.8)	-61.3 (4.0) (0.8)
Change in hours Measure two	57.4	-43.9 (13.7) (2.4)	-42.9 (14.5) (1.2)	-63.1 (5.8) (0.7)	-62.7 (5.4) (0.7)
Percentage change measure one	-2.86%	-3.02% (0.0020) (0.0010)	-2.16% (0.0070) (0.0007)	-3.09% (0.0023) (0.0004)	-3.05% (0.0020) (0.0004)
Percentage change measure two	-2.86%	-2.21% (0.0066) (0.0012)	-2.14% (0.72) (0.0006)	-3.14% (0.0029) (0.0003)	-3.12% (0.0026) (0.0003)

Table 17 Continuous dgp, linear supply function, random preferences, measurement error with standard deviation 200

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1989.5 (19.2) (3.8)	2030.0 (22.0) (3.3)	2009.2 (3.7) (3.6)	2009.5 (3.8) (3.6)
Percentage Bias	0	-0.94% (0.0019)	1.08% (0.0017)	0.04% (0.0018)	0.06% (0.0018)
Hours tax system 5	1951.0	1945.4 (6.9) (4.0)	1988.1 (37.2) (3.1)	1946.1 (5.9) (3.4)	1946.8 (5.3) (3.4)
Percentage Bias	0	-0.29% (0.0021)	1.90% (0.0016)	-0.25% (0.0017)	-0.21% (0.0017)
Change in hours Measure one	57.4	-62.7 (6.2) (3.2)	-20.2 (37.2) (1.7)	-61.9 (4.7) (1.3)	-61.2 (4.1) (1.3)
Change in hours Measure two	57.4	-44.2 (13.7) (3.6)	-42.0 (15.4) (1.2)	-63.1 (5.8) (1.1)	-62.7 (5.4) (1.1)
Percentage change measure one	-2.86%	-3.12% (0.0031) (0.0016)	-1.019% (0.0185) (0.0008)	-3.08% (0.0024) (0.0006)	-3.05% (0.0020) (0.0006)
Percentage change measure two	-2.86%	-2.22% (0.0066) (0.0018)	-2.06% (0.0079) (0.0006)	-3.14% (0.0029) (0.0005)	-3.12% (0.0027) (0.0005)

Table 18 Continuous dgp, linear supply function, random preferences, measurement error with standard deviation 400

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1980.9 (27.9) (5.6)	2086.9 (78.6) (3.9)	2009.0 (4.8) (4.8)	2009.3 (4.8) (4.8)
Percentage Bias	0	-1.36% (0.0028)	3.91% (0.0020)	0.04% (0.0024)	0.05% (0.0024)
Hours tax system 5	1951.0	1937.3 (14.9) (6.1)	2043.4 (92.5) (3.9)	1946.0 (6.6) (4.4)	1946.6 (6.0) (4.3)
Percentage Bias	0	-0.70% (0.0031)	4.74% (0.0020)	-0.26% (0.0022)	-0.22% (0.0022)
Change in hours Measure one	-57.4	-70.5 (14.1) (5.0)	35.5 (93.0) (4.1)	-61.9 (5.1) (2.3)	-61.2 (4.4) (2.2)
Change in hours Measure two	-57.4	-43.6 (15.4) (7.0)	-43.5 (13.9) (1.4)	-63.1 (6.0) (1.8)	-62.6 (5.5) (1.8)
Percentage change measure one	-2.86%	-3.51% (0.0070) (0.0025)	1.77% (0.0463) (0.0021)	-3.08% (0.0025) (0.0011)	-3.05% (0.0022) (0.0011)
Percentage change measure two	-2.86%	-2.20% (0.0074) (0.0035)	-2.08% (0.0077) (0.0007)	-3.14% (0.0030) (0.0009)	-3.12% (0.0027) (0.0008)

Table 19 Continuous dgp, linear supply function, random preferences, measurement error with standard deviation 600

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1960.7 (48.2) (7.5)	2159.3 (151.1) (5.2)	2008.9 (6.2) (6.2)	2009.1 (6.2) (6.1)
Percentage bias	0	-2.37% (0.0037)	7.52% (0.0026)	0.03% (0.0031)	0.04% (0.0031)
Hours tax system 5	1951.0	1918.1 (33.7) (7.2)	2114.8 (163.9) (5.3)	1945.8 (7.5) (5.5)	1946.4 (7.0) (5.4)
Percentage bias	0	-1.69% (0.0037)	8.40% (0.0027)	-0.26% (0.0028)	-0.23% (0.0028)
Change in hours Measure one	-57.4	-89.6 (32.9) (6.8)	107.2 (164.6) (6.4)	-61.9 (5.6) (3.3)	-61.2 (5.0) (3.2)
Change in hours Measure two	-57.4	-42.7 (17.4) (9.3)	-44.5 (12.9) (1.4)	-63.1 (6.3) (2.6)	-62.6 (5.8) (2.5)
Percentage change measure one	-2.86%	-4.46% (0.0164) (0.0033)	5.34% (0.0820) (0.0033)	-3.08% (0.0028) (0.0016)	-3.05% (0.0025) (0.0016)
Percentage change measure two	-2.86%	-2.17% (0.0083) (0.0047)	-2.06% (0.0080) (0.0007)	-3.14% (0.0031) (0.0013)	-3.12% (0.0029) (0.0012)

Table 20 Continuous dgp, linear supply function, random preferences, measurement error with standard deviation 800

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1928.9 (79.9) (8.7)	2354.6 (346.4) (9.9)	2008.7 (7.7) (7.7)	2008.9 (7.7) (7.7)
Percentage bias	0	-3.95% (0.0043)	17.24% (0.0049)	0.02% (0.0038)	0.03% (0.0038)
Hours tax system 5	1951.0	1890.9 (60.6) (8.2)	2301.2 (350.4) (10.3)	1945.7 (8.6) (6.8)	1946.2 (8.1) (6.6)
Percentage bias	0	-3.08% (0.0042)	17.95% (0.0053)	-0.27% (0.0035)	-0.24% (0.0034)
Change in hours Measure one	-57.4	-116.6 (59.9) (9.1)	293.8 (351.3) (10.2)	-61.8 (6.2) (4.4)	-61.2 (5.7) (4.3)
Change in hours Measure two	-57.4	-38.0 (22.1) (10.8)	-53.4 (4.5) (2.1)	-63.0 (6.6) (3.4)	-62.7 (6.3) (3.4)
Percentage change measure one	-2.86%	-5.81% (0.0298) (0.0044)	14.63% (0.1750) (0.0053)	-3.08% (0.0031) (0.0021)	-3.05% (0.0028) (0.0021)
Percentage change measure two	-2.86%	-1.97% (0.0104) (0.0055)	-2.27% (0.0060) (0.0009)	-3.14% (0.0033) (0.0017)	-3.12% (0.0031) (0.0016)

Table 21 Continuous dgp, linear supply function, random preferences, measurement error with standard deviation 1000

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2008.3	1893.0 (115.7) (8.6)	2894.9 (886.7) (18.0)	2008.6 (9.4) (9.4)	2008.7 (9.3) (9.3)
Percentage Bias	0	-5.74% (0.0043)	44.14% (0.0089)	0.01% (0.0047)	0.02% (0.0046)
Hours tax system 5	1951.0	1859.9 (91.6) (10.2)	2858.6 (907.9) (19.7)	1945.5 (9.8) (8.2)	1946.1 (9.3) (8.0)
Percentage Bias	0	-4.67% (0.0052)	46.52% (0.0101)	-0.28% (0.0042)	-0.25% (0.0041)
Change in hours Measure one	-57.4	-147.3 (90.9) (12.6)	851.3 (909.0) (21.6)	-61.7 (7.0) (5.4)	-61.2 (6.6) (5.4)
Change in hours Measure two	-57.4	-33.0 (27.2) (12.2)	-36.3 (21.5) (4.2)	-63.0 (7.1) (4.2)	-62.7 (6.7) (4.2)
Percentage change measure one	-2.86%	-7.34% (0.0452) (0.0061)	42.42% (0.4529) (0.0118)	-3.08% (0.0034) (0.0026)	-3.05% (0.0032) (0.0026)
Percentage change measure two	-2.86%	-1.74% (0.0128) (0.0064)	-1.25% (0.0161) (0.0015)	-3.14% (0.0035) (0.0020)	-3.12% (0.0033) (0.0020)

A third utility function

In the previous sections I have presented simulations with two different utility functions. The first one used was the translog utility function, and it was used both to generate data and to construct the likelihood function. The second utility function used was the one given by eq. (5). When the translog utility function was used, and the data generating process agreed completely with how the likelihood function was constructed, the DC method performed very well. When data was generated with another utility function then the one used for construction of the likelihood function the performance of the DC method was less good. To guard against the possibility that the utility function given by eq. (5) in some way biases results against the DC method and favors the BN method I have also done simulations with a third utility function, much used in the taxable income literature. It is the utility function that generates the so-called log-log taxable income function used by, e.g., Gruber and Saez (2002) and Blomquist and Selin (2010). If this utility function is maximized subject to a linear budget constraint with slope w and intercept y it yields the labor supply function

$$h = kw^{\alpha} y^{\beta} \qquad (6)$$

where k is a constant. This labor supply (utility) function was also used by Burtless and Hausman (1978). One reason why the log log version of this function has been so popular in the taxable income literature is that the parameter α gives the taxable income elasticity if the budget constraint is linear.

In the simulations I use the following values for the parameters. The values for k are drawings from a uniform distribution between 0.8 and 1.0. The mean of the drawings is 0.9. The values of α are drawings from a truncated normal distribution, the lower truncation being 0.28 and the upper 0.3. The underlying normal distribution has mean 0. 28 and standard deviation 0.015. The mean of the truncated distribution is 0.29. The values for β are drawings from a uniform distribution between 0.16 and 0. The mean of the drawings is -0.081. To avoid too many tables I only report results with no measurement errors and with additive measurement errors with standard deviation 400.

Table 22 Continuous dgp, supply function $h = kw^{\alpha} y^{\beta}$, random preferences, no measurement errors

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	2000.3 (11.1) (6.8)	2100.4 (91.4) (5.9)	2009.7 (5.0) (5.0)	2009.9 (5.0) (5.0)
Percentage error	0%	-0.44% (0.0034)	4.54% (0.0029)	0.03% (0.0025)	0.03% (0.0025)
Hours tax system 5	1966.0	1956.5 (11.4) (6.5)	2055.0 (89.3) (5.8)	1962.4 (6.1) (5.0)	1963.3 (5.6) (4.9)
Percentage error	0%	-0.48% (0.0033)	4.53% (0.0030)	-0.18% (0.0025)	-0.14% (0.0025)
Change in hours Measure one	-43.2	-52.0 (9.9) (4.5)	+46.5 (89.8) (2.0)	-46.2 (3.5) (1.8)	-45.2 (2.7) (1.8)
Change in hours Measure two	-43.2	-43.8 (6.1) (6.1)	-45.3 (2.5) (1.3)	-47.4 (4.5) (1.6)	-46.6 (3.7) (1.6)
Percentage change measure one	-2.15%	-2.59% (0.0049) (0.0022)	+2.32% (0.0047) (0.0009)	-2.30% (0.0017) (0.0009)	-2.25% (0.0014) (0.0009)
Percentage change measure two	-2.15%	-2.19% (0.0030) (0.0030)	-2.16% (0.0006) (0.0006)	-2.36% (0.0022) (0.0008)	-2.32% (0.0018) (0.0007)

The dgp:s used to generate the data for tables 16 and 22 are similar in the respect that they both have heterogeneity in preferences and no measurement errors, but they use different labor supply functions. If we compare the results in tables 16 and 22, we see that DC1 in table 22 has a smaller negative bias in the prediction for hours of work for tax system 4, but a larger negative bias for tax system 5, than in table 16. The negative bias in table 22 is of the same magnitude for both tax systems, so the prediction for the change in hours of work using Measure 2 is quite good. In table 16 it was Measure 1 that gave the best prediction of the change in hours of work. Of course, we have no clue of whether eq. (5), which was used to generate data for table 16 or eq. (6), which was used to generate data for table 22, best represent "reality". Hence, the simulations give no clear indication of whether one should use Measure 1 or Measure 2 to predict the change in hours of work if one uses DC1. For DC2 the prediction of hours of work are pretty bad for both tax systems 4 and 5, when data are generated by the labor supply function given by eq. (6). However, the bias is almost of exactly the same size, so the prediction of the change in hours of work is very precise if Measure 2 is used. This is of little interest, since, overall it is DC1 that shows the best performance of the DC estimators. The BN method performs slightly better when data are generated by eq. (6) than eq. (5), but the difference is not large. The Lasso version is slightly better than the OLS version.

Table 23 Continuous dgp, supply function $h = kw^{\alpha} y^{\beta}$, random preferences, measurement errors with standard deviation 400

	True	DC1, with GEV errors	DC2, no GEV errors	BN-OLS	BN-Lasso
Hours tax system 4	2009.2	1979.2 (30.8) (7.1)	2153.0 (144.0) (6.5)	2009.5 (6.2) (6.2)	2009.6 (6.3) (6.3)
Percentage error	0%	-1.49% (0.0035)	7.16% (0.0032)	0.02% (0.0031)	0.02% (0.0031)
Hours tax system 5	1966.0	1939.2 (27.9) (7.8)	2108.2 (142.4) (6.9)	1962.4 (6.8) (5.9)	1963.2 (6.5) (5.9)
Percentage error	0%	-1.36% (0.0040)	7.23% (0.0035)	-0.18% (0.0030)	-0.14% (0.0030)
Change in hours Measure one	-43.2	-69.4 (27.1) (6.8)	+99.6 (142.9) (4.6)	-46.2 (4.0) (2.7)	-45.4 (3.4) (2.7)
Change in hours Measure two	-43.2	-40.0 (8.9) (8.3)	-44.8 (2.4) (1.8)	-47.1 (4.5) (2.2)	-46.3 (3.8) (2.2)
Percentage change measure one	-2.15%	-3.46% (0.0135) (0.0033)	+4.96% (0.0711) (0.0024)	-2.30% (0.0020) (0.0013)	-2.26% (0.0017) (0.0013)
Percentage change measure two	-2.15%	-2.02% (0.0043) (0.0042)	-2.08% (0.0011) (0.0008)	-2.34% (0.0022) (0.0011)	-2.31% (0.0019) (0.0011)

Results where we add measurement errors are shown in table 23. The measurement errors are drawn from a normal distribution with mean zero and standard deviation 400. The DC predictions of hours of work for tax systems 4 and 5 are worse than those presented in table 18, especially for DC2. However, since for both DC estimators, the negative bias in the predictions are of similar magnitude, the prediction of the change in hours of work using Measure 2 is better in table 23 than in table 18. For the BN method the addition of a measurement error does not affect the bias, but increase the RMSE, which is the same pattern as for the earlier utility functions. As for the earlier simulations the results for the Lasso version and the OLS versions are similar, but the Lasso version slightly better. Using the Lasso version of the BN method and Measure 1 gives the best prediction of the change in hours of work.

5. Summary

In this article I have used Monte Carlo simulations to study how well the DC and BN methods can predict the effect of a tax reform. In section 3 of the paper I use a discrete choice dgp where individuals can choose between 11 points of hours of work spaced evenly between 0 and 3000 hours. This dgp is in accordance with the dgp assumed in the discrete choice literature. In section 4 I use a dgp where individuals can choose freely among all points on a continuous piece-wise linear budget constraint. For the DC method two alternatives are studied, one where GEV errors are added when doing the prediction, DC1, and another where GEV errors are not added, DC2. For the BN method I try both a version where Lasso coefficients are used in the prediction, BN-Lasso, and another where OLS coefficients are used, BN-OLS. The OLS is run on the same variables as picked by the LASSO. The setup is that data are generated for 5 different tax systems. Data for four of the tax systems are used for estimation. The estimated labor supply functions are then used to predict the effect of a tax reform that takes us from tax system 4 to tax system 5. The estimated functions are used to predict the average hours of work for the fourth and a fifth tax systems. Data from the fourth tax system is part of the data set used in the estimation, so the prediction for the fourth tax system is a within sample prediction, whereas the prediction for the fifth tax system is an out of sample prediction. The change in average hours can be calculated either as the difference between predicted hours for tax system 5 minus the average hours of the data from tax system 4 that was used in the estimation, Measure 1, or as the difference between the predicted hours of work for tax systems 4 and 5, Measure 2.

An overall conclusion of the simulations is that DC1 is more reliable than DC2 and that BN-Lasso performs slightly better than BN-OLS. I therefore confine the discussion in this summary to DC1 and BN-Lasso. Moreover, the simulations show that results are very similar for the discrete dgp and the continuous dgp. In the following when numbers are used they are from the discrete dgp simulations.

In the first three simulations the dgp aligns precisely with the assumptions used when deriving the DC1 likelihood function. DC1 also predicts in the same way as data are generated. In the first simulation the GEV errors are so small, in relation to the utilities from the deterministic part of the utility function, that they do not affect the distribution of hours of work. For this simulation we find that DC1 performs very well. BN-Lasso is not bad, but not as good as DC1. Somewhat surprisingly, for the third simulation, where the GEV errors play a more important role for the distribution of hours of work, in terms of RMSE, the BN-Lasso performs better than DC1.

We successively make changes in the dgp. The first change is to use another utility function than the one used in the DC1 likelihood function and not to add GEV errors when generating data. This implies that the DC1 likelihood function will be misspecified. We let preferences be homogenous, so that all individuals have the same preferences. Results for this dgp are mixed and DC1 is not a clear winner as in table 1. BN-Lasso does a better job in predicting hours of work for tax system 4, whereas DC1 does a better job in predicting hours of work for tax system 5. Using DC1 and Measure 2 gives the best prediction of the change in hours of work.

The next two changes in the dgp is to make preferences heterogeneous and add optimization/measurement errors. These changes hopefully give a dgp that better mirror how real data are generated. We start by making preferences heterogenous while keeping the assumption of no measurement errors. As compared to when preferences are homogenous both DC1 and BN-Lasso perform better. The BN-Lasso out performs DC1 both in terms of bias and RMSE whatever prediction we look at. The best prediction of the change in hours of work obtains when BN-Lasso and Measure 1 is used for the prediction.

Further simulations explore how the performance of the estimators are affected if there is an additive normally distributed optimization/measurement error in hours of work. In these simulations the standard deviation of the optimization/measurement error increase from zero up to 1000 hours. When there are no optimization/measurement errors, DC1 has a downward percentage bias of -0.80% for the prediction of average hours of work for tax system 4 and -0.24% for tax system 5, the RMSE is 17.8 for the prediction for tax system 4 and 5.9 for tax system 5. When measurement errors with standard deviation 200 are added to hours of work these downward biases increases to -0.97% and -0.34% and the RMSE to 20.0 and 7.8. As the standard deviation of the errors increases the downward bias increases successively to reach a downward percentage bias of -5.75% for average hours of work for tax system 4 and -4.65% for tax system 5 and with RMSE 115.8 and 91.3 respectively, when the standard deviation for the optimization/measurement error is 1000. For BN-Lasso, the percentage bias of the prediction for the average hours of work for tax system 4 stays pretty much the same as we increase the standard deviation of the measurement error and is around 0 to 0.05%. The percentage bias for the prediction for tax system 5 also stays roughly the same and is around 0.25% to 0.30%. However, as we increase the dispersion of the measurement errors, the RMSE increases, this increase is because the standard deviation of the estimates increases. The RMSE for the prediction of average hours of work for tax system 4 increases from 4.2 when the standard deviation is 200 to 10 when the standard deviation is 1000 and for tax system 5 from 6.2 to 10.9. In all these simulations using BN-Lasso and Measure 1 gives the best prediction of the change in hours of work.

To guard against the possibility that the two utility (labor supply) functions used for the majority of the simulations in some way biases the results against the DC method and favors the BN method I have also done simulations with a third utility function; a utility (labor supply) function much used in the taxable income literature. The results from those simulations are in line with the results from the other utility (labor supply) functions.

It is sometimes said that one of the advantages of the DC method is that it can easily handle nonconvex budget constraints. Computationally that is correct. However, two simulations with heterogenous preferences and nonconvex budget constraints indicate that DC1 performs less well in terms of bias and RMSE when budget constraints are nonconvex. As compared to when the budget constraints are convex, and there are no measurement errors, the bias and RMSE becomes larger for DC1 when budget constraints are nonconvex. For BN-Lasso the bias for hours of work for tax system 4 is about the same in absolute value as when the budget constraints are convex. For hours of work for tax system 5 the bias is lower. The RMSE for the BN-Lasso predictions of the average hours of work is lower for both tax systems. If there are measurement errors with a standard deviation of 400 the bias for the DC1 prediction of average ours of work for tax system 4 is -1.94% and -0.91% for tax system 5. For BN-Lasso the corresponding biases are -0.03 and -0.11 percent. The RMSE for the DC1 prediction of hours of work for tax system 4 is 39.3 and for tax system five 18.8. The corresponding RMSE for BN-Lasso are 4.7 and 6.2.

All simulations except one, are done with a sample size of 40,000. To study how the properties, depend on sample size I have done a simulation with a sample size of 2,000 and optimization/measurement errors with a standard deviation of 400. The percentage bias in the DC1 predictions of average hours of work increase from -1.39% and -0.76% for tax systems 4 and 5 respectively, when the sample size is 80,000, to -1.52% and -0.87% when the sample size is 2,000. The corresponding figures for BN-Lasso is an increase from 0% and -.26% to -0.09% and -0.33%. For DC1 the RMSE increase from 28.5 for average hours for tax system 4 and 16.0 for tax system 5 to 40.8 and 33.0. The corresponding numbers for BN-Lasso are from 5.4 to 22.0 for tax system 4 and 7.1 to 21.3 for tax system 5. As when the sample size is 80,000, the BN-Lasso performs better than DC1 both in terms of bias and RMSE when the sample size is 2,000.

Except for simulations where the dgp aligns exactly with the DC1 likelihood function, BN-Lasso dominates DC1 both in terms of predicting average hours of work for a tax system and the change in average hours of work due to tax reform.

Appendix A. More on the Discrete Choice Method

Let there be J alternative values for hours of work indexed by j and N individuals indexed by *i*. To each value of hours of work, h_i^j, there is a value of consumption, c_i^j. The utility of alternative *j* for individual *i* is given by

$$v\left(c_i^j, h_i^j\right) = u\left(c_i^j, h_i^j\right) + \varepsilon_i^j \qquad j=1,\ldots,J \text{ and } i=1,\ldots,N$$

where $u\left(c_i^k, h_i^i\right)$ is a weakly quasi-concave function and the ε_i^j are drawings from the GEV 1 extreme value distribution. Individuals chose the point which is the maximal element of the $v\left(c_i^j, h_i^j\right)$. Then the probability that individual i chooses alternative k is given by

$$p_i^k = \frac{\exp(U_i^k)}{\sum_{j=1}^{J} \exp\left(U_{ij}\right)} \qquad \text{(A1)}$$

The likelihood of a sample is given by

$\mathcal{L} = \Pi_{i=1}^{N} p_i^k$ and the log likelihood by log $\mathcal{L} = \sum_{i=1}^{N} \log p_i^k = \sum_{i=1}^{N} U_i^k - \sum_{i=1}^{N} \log\left(\sum_{j=1}^{K} \exp\left(U_i^j\right)\right)$

Alternatively, we can write the log likelihood function as

$$\mathcal{L} = \sum_{i=1}^{N} \log p_i^k = \sum_{i=1}^{N} \log\left(\frac{\exp(U_i^k)}{\sum_{j=1}^{K} \exp\left(U_i^j\right)} \right).$$

Accounting for heterogeneity in preferences when constructing the DC likelihood function

Heterogeneity in individuals' preferences that can be represented by observable characteristics can easily be incorporated both in the BN and the DC methods; that will not be studied here. Earlier studies have represented unobservable heterogeneity in the utility function in two ways, the random coefficient model (Van Soest, 1995) and the latent class model (Hoynes, 1996). As the random coefficient model has become the standard way to allow for unobservable heterogeneity, it is this method I discuss here.

Van Soest (1995) uses a translog utility function in the estimation of a household utility function with hours of leisure of wife and husband respectively and total after tax income (consumption) as arguments. He allows the coefficients multiplying the log of wife's leisure hours and the log of husband's leisure hours to be random variables. These coefficients are represented as a fixed coefficient, same for all, plus a normally distributed random variable with mean zero and a standard deviation to be estimated from the data. Introducing random coefficients complicates the formulation of the likelihood function.

Let θ denote a vector of parameters, where one or more of the parameters is a random variable. Let x be a vector of independent variables and $v\left(\left(x:\theta\right)\right)$ the utility function. Let x_i^k

be the alternative that the $i:th$ individual chooses and x_i^j be the vector of independent variables for the $j:th$ alternative. Then the contribution to the likelihood function for the $i:th$ observation would be $\int \frac{\exp(v\left(x_i^k:\theta\right)}{\sum \exp\left(v\left(x_i^j:\theta\right)\right)} f\left(\theta\right) d\theta$. Normally this expression cannot be written in a closed analytic form, but has to be evaluated numerically. The common way to do this is to draw R random numbers from the distribution in question and approximate the integral by the average over the random drawings, i.e., $\frac{1}{R}\sum_{r=1}^{R}\frac{\exp\left(v\left(x_i^k:\theta^r\right)\right)}{\sum_{j=1}^{K}\exp\left(v\left(x_i^j:\theta^r\right)\right)}$. Gong and van Soest (2002) try different values for R and finds that R=20 is enough. Van Soest et al. (2002) come to the same conclusion. The form of the log likelihood function then takes the form

$$\ln(L)=\sum_{i=1}^{N}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\frac{\exp\left(v\left(x_i^k:\theta^r\right)\right)}{\sum_{j=1}^{J}\exp\left(v\left(x_i^j:\theta^r\right)\right)}\right)$$

The estimates of the standard deviation of the random terms obtained in Van Soest (1995) are imprecise. The t-value for the standard error for the random term in the parameter multiplying the log of the husband's leisure hours is 1.18 and the t-value for the standard error for the random term in the parameter multiplying the log of the wife's leisure hours is 0.22. Van Soest et al. (2002), which use a polynomial utility function, introduce a random component in the parameter multiplying the linear term for hours of work. This random term is assumed to be normally distributed with mean zero and variance σ_{rp}^2. The variance is estimated to zero and the t-value of this estimate is also zero. Hence, in both these studies by Soest it seems hard to estimate the variance of the random term. Löffler et al. (2018) studies different specifications of the discrete choice model. With respect to accounting for unobservable preference heterogeneity, they conclude that accounting for unobserved heterogeneity yields little value added in term of statistical fit.

To judge how important it is to account for heterogeneity in the way suggested in Van Soest (1995), and used by many others, I will perform two simulations, one where the dgp is discrete and another where the dgp is continuous. In our simulations, we use a simpler model with just leisure hours and consumption in the utility function, so in our context it would be the parameter multiplying $\ln\left(L\right)$ that would be a random variable. In the construction of the likelihood function it is assumed that the parameter multiplying $\ln(L)$ is a constant, the same for everyone, plus a normally distributed term with mean zero and a standard deviation to be estimated from the data.

When constructing the data I use the utility function given by eq. (2) and all the preference parameters vary across individuals. Individuls' s_i are drawings from a normal distribution with mean 1.25 and standard deviation 0.12. The wage rate coefficients α_i are drawings from a truncated normal distribution; the untruncated distribution having mean 0.014 and standard deviation 0.0015, the lower truncation being 0.008 and the upper truncation being 0.020. The mean of the truncated distribution is 0.014 and the standard deviation 0.0015. The nonlabor income coefficients β_i are drawings from a truncated normal distribution; the untruncated distribution having mean 0.2 and standard deviation 0.009, the lower truncation being -0.002 and the upper zero. The mean of the truncated distribution is -0.00039 and the standard deviation 0.00037.

I use samples of ten million individuals. The estimates of the standard deviation for the random term is zero for both the discrete and continuous dgp. Hence, it is of no value to use the more complicated likelihood function that takes account of unobserved heterogeneity in the way suggested in Van Soest (1995). This is essentially the same finding as the findings in Van Soest (1995) and Van Soest et al. (2002). In the main simulations I have therefore not used the likelihood function that accounts for heterogeneity.

Appendix B: The BN method

When hours of work are determined by maximization of a quasi-concave utility function subject to a linear budget constraint with intercept y and slope w, we can write hours of work as a function of y and w; $h = h(y, w)$. Since there are only two arguments in this function one can estimate $h(y, w)$ nonparametrically. The basic idea of the BN-method is to think of the choice of hours of work as a function of the entire budget set also in the case where the budget set is nonlinear and consists of many piece-wise linear segments. It is a method to nonparametrically estimate the conditional mean of hours given a piece-wise linear budget set. That is, if h_i is the hours of the i^{th} individual and x_i represents the budget set the method shows how to nonparametrically estimate

$$E[h_i \mid x_i] = \bar{h}(x_i) \qquad \text{(B1)}$$

Before going into the details of how to nonparametrically estimate $\bar{h}(x)$ we note that the BN-method, which is a regression method, in contrast to the DC method, can handle both additive and multiplicative optimization/measurement errors. Let $h_i = h(x_i)$ be desired hours of work for individual i if the budget constraint is x_i. Let observed hours of work be $\tilde{h}_i = \varepsilon_{i1} h(x_i) + \varepsilon_{i2}$, where ε_{i1} is a multiplicative optimization/measurement error and ε_{i2} an additive optimization/measurement error. If the error terms and the budget constraint x_i are stochastically independent, and $E(\varepsilon_{i1}) = 1$ and $E(\varepsilon_{i2}) = 0$, then $E(\tilde{h}_i \mid x_i) = E(h_i \mid x_i) = \bar{h}(x_i)$.

A piecewise linear budget constraint is completely described by

$$x = (J, y_1, \ldots y_J, w_1, \ldots w_J, l_1, \ldots l_{J-1})$$

where J is the number of segments, y_j and w_j the intercept and slope of the j^{th} segment and l_j the endpoint between segments j and $j+1$. If the budget frontier is continuous, then $l_j = (y_j - y_{j+1})/(w_{j+1} - w_j)$ and the l_j can be dropped from x, reducing the dimension of x somewhat. Still, x will be of a large dimension and it would be hard to estimate $\bar{h}(x)$ nonparametrically if we do not have some way to reduce the dimensionality. Luckily, if we assume data are generated by utility maximization subject to a convex budget set, then the dimensionality of the nonparametric estimation can be reduced dramatically.

Let $\tilde{\pi}(y, w, v)$ be the solution to utility maximization with a quasi-concave utility function subject to a linear budget constraint with slope w and intercept y. Following Blomquist and Newey (2002) we assume v is a scalar unobserved variable that represents individual heterogeneity and that $\tilde{\pi}(y, w, v)$ is increasing in v. (We make the assumption that v is a scalar to ease the exposition. It is shown in Blomquist et al. (2014) and Blomquist et al (2023) that the results shown below also holds if v is a many dimensional vector). For sufficiently negative v it is possible that $\tilde{\pi}(y, w, v)$ is negative. Since negative hours of work are impossible we define $\pi(y, w, v) = \max\{0, \tilde{\pi}(y, w, v)\}$.

For ease of exposition we will derive an expression for $E[h_i \mid x_i] = \bar{h}(x_i)$ for a piece-wise linear budget constraint with just two segments and one kink. It is straightforward to extend the derived expression to budget constraints with any number of segments and kinks. If we have a first segment with slope w_1 and intercept y_1 and start from a large negative value of v, the desired hours of work will be zero, at some value of v, $v_0 = v(y_1, w_1, 0)$ there will be a tangency solution between an indifference curve of the utility function and the first segment at zero hours of work, as v increases further the tangency point will move up the first segment until we come to a v value such that there is a tangency at l_1, $v = v(y_1, w_1, l_1)$. If we increase v further there will be a tangency point on the extended first segment. However, this tangency point is not within the budget set, so the point cannot be chosen. Instead there will be an interval of v for which the chosen hours of work will be at l_1, we denote this interval $[\underline{v}(y_1,w_1,l_1),\overline{v}(y_2,w_2,l_1)]$. The lower value of the interval is the value of v for which there is a tangency between an indifference curve and the first segment at the point l_1. The upper value of the interval is the value of v such that there is a tangency between an indifference curve and the second segment at l_1. As v increases further the tangency point will move upwards on segment 2. Presumably there is a value of v such that the pdf is zero at that value, and zero for all higher values of v.

The function $\pi(y,w,v)$ is increasing in v. This means that we can invert the function. For each value of $h=\pi(y,w,v)$ there is a unique value of v. We write the inverse function as $v=\pi^{-1}(y,w,h)$ We can therefore write $\underline{v}(y_1,w_1,l_1)=\pi^{-1}(y_1,w_1,l_1)$

and $\overline{v}(y_2,w_2,l_1)=\pi^{-1}(y_2,w_2,l_1)$.

We now know there is a first interval of v for which desired hours are on the first segment, a second interval of v for which desired hours will be at the kink point, and a third interval of v for which desired hours will be on the second segment. Denoting the pdf of v by $g(v)$ we can therefore write the expected hours of work as:

$$E(h\mid x)=\int_{-\infty}^{\pi^{-1}(y_1,w_1,l_1)}\pi(y_1,w_1,v)g(v)\,dv+\int_{\pi^{-1}(y_1,w_1,l_1)}^{\pi^{-1}(y_2,w_2,l_1)}l_1g(v)\,dv+\int_{\pi^{-1}(y_2,w_2,l_1)}^{\infty}\pi(y_2,w_2,v)g(v)\,dv \quad (B2)$$

The first integral gives the contribution to the expected value for v such that desired hours is on the first segment, the second integral gives the contribution to the expected value for values of v such that desired hours is at the kink, and the last integral gives the contribution to the expected value for v such that desired hours are on the second segment.

We can rewrite the second integral of (B2) as:

$$\int_{\pi^{-1}(y_1,w_1,l_1)}^{\pi^{-1}(y_2,w_2,l_1)}l_1g(v)\,dv=\int_{-\infty}^{\pi^{-1}(y_2,w_2,l_1)}l_1g(v)\,dv-\int_{-\infty}^{\pi^{-1}(y_1,w_1,l_1)}l_1g(v)\,dv.$$

and the last integral as

$$\int_{\pi^{-1}(y_2,w_2,l_1)}^{\infty}\pi(y_2,w_2,v)g(v)\,dv = \int_{-\infty}^{\infty}\pi(y_2,w_2,v)g(v)\,dv-\int_{-\infty}^{\pi^{-1}(y_2,w_2,l_1)}\pi(y_2,w_2,v)g(v)\,dv \text{ -}$$

Using these results we can rewrite (B2) as:

$$E(h\mid x)=\int_{-\infty}^{\infty}\pi(y_2,w_2,v)g(v)\,dv+\int_{-\infty}^{\pi^{-1}(y_1,w_1,l_1)}\left[\pi(y_1,w_1,v)-l_1\right]g(v)\,dv$$

$$-\int_{-\infty}^{\pi^{-1}(y_2,w_2,l_1)}\left[\pi\left(y_2,w_2,v\right)-l_1\right]g\left(v\right)dv \qquad \text{(B3)}$$

Ii is straightforward to generalize this result so that if there were J segments the expression for expected hours of work would be

$$E\left(h\,|\,x\right)=\int_{-\infty}^{\infty}\pi\left(y_J,w_J,v\right)g\left(v\right)dv+$$

$$\sum_{j=1}^{J-1}\left[\int_{-\infty}^{\pi^{-1}(y_j,w_j,l_j)}\left[\pi\left(y_j,w_j,v\right)-l_j\right]g\left(v\right)dv-\int_{-\infty}^{\pi^{-1}(y_{j+1},w_{1j+1},l_j)}\left[\pi\left(y_{j+1},w_{j+1},v\right)-l_j\right]g\left(v\right)dv\right].$$

To simplify notation we define the function

$\bar{\pi}\left(y,w\right)=\int_{-\infty}^{\infty}\pi\left(y,w,v\right)g\left(v\right)dv$, i.e., the expected hours of work for a linear budget constraint with intercept y and slope w, and the function

$$\mu\left(y,w,l\right)=\int_{-\infty}^{\pi^{-1}(y,w,l)}\left[\pi\left(y,w,v\right)-l\right]g\left(v\right)dv$$

We can then rewrite (B3) as

$$E\left(h\,|\,x\right)=\bar{\pi}\left(y_J,w_J\right)+\sum_{j=1}^{J-1}\left[\mu\left(y_j,w_j,l_j\right)-\mu\left(y_{j+1},w_{j+1},l_j\right)\right] \qquad \text{(B4)}$$

The first term, $\bar{\pi}\left(y_J,w_J\right)$, is the expected hours of work for the extended last segment. We can regard the second term as a correction term due to the nonlinearity of the budget constraint.

Instead of having to nonparametrically estimate a function with 2J elements, we can estimate one function $\bar{\pi}\left(y_J,w_J\right)$ with two arguments and another function $\mu\left(y_j,w_j,l_j\right)$ with three arguments. This means that nonparametric estimation is feasible. Blomquist and Newey (2002) also show how minor nonconvexities of the budget constraint can be handled. However, I will not show that here.

Following Blomquist and Newey (2002), in the simulations I will approximate $\bar{\pi}\left(y_J,w_J\right)$ and $\mu\left(y_j,w_j,l_j\right)$ by power functions and estimate (B4) by series estimation. I approximate $\bar{\pi}\left(y_J,w_J\right)$ as:

$\bar{\pi}(y_J, w_J) \approx 1 + \sum_{k=1}^{K} \gamma_k y_J^{p(k)} w_J^{q(k)}$, where γ_k are parameters to be estimated. Here K will be nine and the terms with highest power will be y_J^3 and w_J^3. I approximate $\mu(y_j, w_j, l_j)$ as $\mu(y_j, w_j, l_j) \approx \sum_{k=1}^{K} \sigma_k l_j^{m(k)} y_j^{p(k)} w_j^{q(k)}$, $p(k) + q(k) \geq 1$, where σ_k are parameters to be estimated.

Given our choice to approximate the functions we can rewrite (B4) in the following way

$$E(h \mid x) = \bar{\pi}(y_J, w_J) + \sum_{j=1}^{J-1} \left[\mu(y_j, w_j, l_j) - \mu(y_{j+1}, w_{j+1}, l_j) \right] \approx$$

$$1 + \sum_{k=1}^{K} \gamma_k y_J^{p(k)} w_J^{q(k)} + \sum_{j=1}^{J-1} \left[\sum_{k=1}^{K} \sigma_k l_j^{m(k)} y_j^{p(k)} w_j^{q(k)} - \sum_{k=1}^{K} \sigma_k l_j^{m(k)} y_{j+1}^{p(k)} w_{j+1}^{q(k)} \right] =$$

$$1 + \sum_{k=1}^{K} \gamma_k y_J^{p(k)} w_J^{q(k)} + \sum_{j=1}^{J-1} \left[\sum_{k=1}^{K} \sigma_k l_j^{m(k)} \left(y_j^{p(k)} w_j^{q(k)} - y_{j+1}^{p(k)} w_{j+1}^{q(k)} \right) \right] =$$

$$1 + \sum_{k=1}^{K} \gamma_k y_J^{p(k)} w_J^{q(k)} + \sum_{k=1}^{K} \sigma_k \sum_{j=1}^{J-1} l_j^{m(k)} \left[y_j^{p(k)} w_j^{q(k)} - y_{j+1}^{p(k)} w_{j+1}^{q(k)} \right] \qquad \text{(B5)}$$

In the following I will denote $\sum_{j=1}^{J-1} l_j^{m(k)} \left[y_j^{p(k)} w_j^{q(k)} - y_{j+1}^{p(k)} w_{j+1}^{q(k)} \right]$ by $l^{m(k)} \Delta y^{p(k)} w^{q(k)}$. To exemplify, suppose $J = 4$, $m(k) = 0$, $p(k) = 1$ and $q(k) = 1$, so that a $l^{m(k)} \Delta y^{p(k)} w^{q(k)} = \Delta yw$, then the $k:th$ regressor in $\sum_{k=1}^{K} \sigma_k \sum_{j=1}^{J-1} l_j^{m(k)} \left[y_j^{p(k)} w_j^{q(k)} - y_{j+1}^{p(k)} w_{j+1}^{q(k)} \right]$ would be the summation $\sum_{j=1}^{3} \left(y_j w_j - y_{j+1} w_{j+1} \right)$. A beauty with this type of terms is that the dimensionality of the estimation problem does not depend on the number of segments and that the number of segments can differ between observations. The approximation would of course be better if we used many approximating terms, i.e. a large value for K. However, in the estimation we know that the bias of our estimates decreases as we increase the number of approximating terms, but the variance increases. In the estimation and simulations, I will use the same approximating terms as those used in Blomquist and Newey (2002). To approximate the function $\bar{\pi}(y_J, w_J)$ we use the terms: $1, y_J, w_J, y_J^2, w_J^2, y_J w_J, y_J^3, w_J^3, y_J^2 w_J, y_J w_J^2$ and to approximate $\sum_{j=1}^{J-1} \left[\mu(y_j, w_j, l_j) - \mu(y_{j+1}, w_{j+1}, l_j) \right]$ we use the terms $\Delta y, \Delta w, l\Delta y, \Delta y^2, \Delta w^2, \Delta yw, l^2 \Delta y, l\Delta y^2, l\Delta w^2, l\Delta yw$. To avoid perfect multicollinearity among the regressors, $l\Delta w$ and $l^2 \Delta w$ are not included.

We see that the assumption of utility maximization implies a very impressive reduction in dimensionality. The expected value only depends on a term with the summation of differences.

A term that is independent of the number of segments. This implies that different individuals can have different number of segments and if we have data from different tax systems it is of no consequence if the number of segments differ between the tax systems.

Trade off between variance and bias

A crucial question is how to choose the number of terms in the approximating function. How many terms to include depends on if we want to use the estimated function for inference or prediction. Here I assume that the estimated function, whether estimated by a DC-method or a BN-method, will be used for prediction. Models with high variance are usually more complex (e.g. higher-order regression polynomials), enabling them to represent a training set more accurately. In the process, however, they may also represent a large noise component in the training set, making their out of sample predictions less accurate – despite their added complexity. In contrast, models with higher bias tend to be relatively simple (low-order or even linear regression polynomials) but may produce lower variance predictions when applied beyond the training set. To select terms to be included I will use the Lasso method. A sequence of different values of lambda are used and a cross-validation measure will be used to choose a value of lambda. In the simulations where the sample size for the estimation is 40,000, the variance (in the training set) will never become large and the effect that the bias goes down with the number of terms leads to a minimum for the cross-validation measure when all variables are included. In the tables I show both predictions using the Lasso coefficients obtained for the lambda value where the cross-validation is minimized and the post-Lasso, i.e. OLS on all the terms.

Further developments of the BN method

Blomquist and Newey (2002) did not use the fact that it is the same pdf $g(v)$ that generates the functions $\bar{\pi}(y,w)=\int_{-\infty}^{\infty}\pi(y,w,v)g(v)dv$ and $\mu(y,w,l)=\int_{-\infty}^{\pi^{-1}(y,w,l)}\left[\pi(y,w,v)-l\right]g(v)dv$. In Blomquist et al (2015) its is shown how to use this information and obtain functions to be estimated that should improve the efficiency of the method. In Blomquist et al. (2015) it is also shown that the formulas derived in Blomquist and Newey (2002) are valid also in the case where v is multidimensional. In Blomquist et al (2023) the method is further developed. To choose the number of terms to be included in the estimation the Lasso procedure is used. However, such a regularization introduces a bias in the estimates. In Blomquist et al. (2023) it is shown how to take account of this by using a double debiased machine learning method. These developments are technically more demanding than the version of the BN method derived in Blomquist and Newey (2002); in this paper I have therefore chosen to use the version of the BN method described in Blomquist and Newey (2002).

References

Aaberge, R., J. Dagsvik and S. Ström, 1995, Labor Supply Responses and Welfare Effects of Tax Reforms, *Scandinavian Journal of Economics* 97, 635-659, 1995.

Aaberge R. and U. Colombino, 2014, Labour Supply Models, in C. O'Donoghue (Ed.), *Handbook of Microsimulation Modelling*, Emerald 2014.

Beffy, M., R. Blundell, A Bozio, G. Laroque and M. To, 2019, "Labour Supply and Taxation with Restricted Choices", *Journal of Econometrics* 211, 16-46.

Belloni, A. and V. Chernozhukov, 2013, ″Least Squares After Model Selection in HighDimensional Sparse Models," *Bernoulli* , 521-547.

Blomquist, S., 1983, "The Effect of Income Taxation on the Labor Supply of Married Men in Sweden", *Journal of Public Economics* 22, 169-197.

Blomquist, S. and U. Hansson-Brusewitz, 1990, "The Effect of Taxes on Male and Female Labor Supply in Sweden", *Journal of Human Resources* 25, 317-357.

Blomquist, S., 1996, "Estimation Methods for Male Labor Supply Functions: How to Take Account of Nonlinear Taxes", *Journal of Econometrics* 70, 383-405.

Blomquist, S., Kumar, A., Liang, C. Y., & Newey, W. (2015): "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," CESifo Working Paper No. 5320.

Blomquist, S., Kumar, A., Liang, C. Y., & Newey, W. (2023): "Nonlinear Budget Set Regressions for the Random Utility Model", NBER Working Paper 31194.

Blomquist, S. and W. Newey (2002): "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica* 70 : 2455–2480.

Blomquist, S. and H. Selin (2010): "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates," *Journal of Public Economics* 94 : 878–889.

Blundell, R. and A. Shephard, 2008, Employment, Hours of Work and the Optimal Taxation of Low-Income Families, *Review of Economic Studies* 79, 481–510.

Burtless, G. and J. Hausman (1978): "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy* 86 : 1103–1130.

Dagsvik, J., 1994, Discrete and continuous choice, max-stable processes and independence from irrelevant attributes, *Econometrica* **62**, 1179-1205.

Duncan, G.J. and D. Hill (1985). "An Investigation of the Extent and Consequences of Measurement Error in Labor Economic Survey Data", *Journal of Labor Economics,* 3, pp. 508-522.

Gruber, J. and E. Saez (2002): “The Elasticity of Taxable Income: Evidence and Implications,” *Journal of Public Economics* 84 :1–32

Kumar, A, 2012, Nonparametric estimation of the impact of taxes on female labor supply, *Journal of Applied Econometrics* 27, 415-439.

Liang, C-Y, Nonparametric structural estimation of labor supply in the presence of censoring, *Journal of Public Economics* 96, 89-103.

Löffler, M., A. Peichl and S. Siegloch, 2018, “The Sensitivity of Structural Labor Supply Estimations to Modelling Assumptions”, IZA Discussion Paper No. 11425

McFadden, D. ,1974, Conditional logit analysis of qualitative choice behavior, Frontiers in Econometrics, New York, Academic Press, p 105-142.

Meyer, B.D. and N. Mittag, 2017, Misclassification in binary choice models, *Journal of Econometrics*, 200, 295-311.

Stern, N. (1986), “On the Specification of Labour Supply Functions,” in R. Blundell and I. Walker (editors) Unemployment, Search and Labour Supply, Cambridge: Cambridge University Press, 143-189.

Tibshirani, R.,1996, Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267-288.

van Soest, A., 1995, Structural models of family labor supply – a discrete choice approach, *Journal of Human Resources* 30(1), 63–88.

van Soest, A., Das, M., & Gong, X, 2002, A structural labour supply model with flexible preferences, *Journal of Econometrics* 107, 345-374.