Analysing Tax-Benefit Reforms in the Netherlands: Using Structural Models and Natural Experiments.

Boer, H.W. de; Jongen, E.L.W.

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12892

Analysing Tax-Benefit Reforms in the Netherlands: Using Structural Models and Natural Experiments

Henk-Wim de Boer
Egbert L.W. Jongen

JANUARY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12892

Analysing Tax-Benefit Reforms in the Netherlands: Using Structural Models and Natural Experiments

Henk-Wim de Boer
CPB Netherlands Bureau for Economic Policy Analysis

Egbert L.W. Jongen
CPB Netherlands Bureau for Economic Policy Analysis, Leiden University and IZA

JANUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Analysing Tax-Benefit Reforms in the Netherlands: Using Structural Models and Natural Experiments*

We combine the strengths of structural models and natural experiments in the analysis of tax-benefit reforms in the Netherlands. First we estimate structural discrete-choice models for labour supply. Next we simulate key past reforms and compare the predictions of the structural model with the outcomes of quasi-experimental studies. The structural model predicts the treatment effects well. The structural model then allows us to conduct counterfactual policy analysis. Policies targeted at working mothers with young children generate the largest labour supply responses, but generate little additional government revenue. Introducing a at tax, basic income or joint taxation is not effective.

JEL Classification: C25, C52, H31, J22

Keywords: structural models, natural experiments, tax-benefit reform, Netherlands

Corresponding author:
Egbert L.W. Jongen
CPB Netherlands Bureau for Economic Policy Analysis
Bezuidenhoutseweg 30
2594 AV, The Hague
The Netherlands
E-mail: e.l.w.jongen@cpb.nl

* We are grateful to Peter Dekker for his assistance in the construction of the simulation model (MICSIM). Furthermore, we have benefitted from comments and suggestions by Rolf Aaberge, Olivier Bargain, Hans Bloemen, Richard Blundell, Stefan Boeters, Casper van Ewijk, Rob Euwals, Laura van Geest, Jan Kabatek, Chris van Klaveren, Tom Kornstad, Arjan Lejour, Maarten Lindeboom, Ruud de Mooij, Andreas Peichl, Arthur van Soest, Coen Teulings, Thor Thoresen, Trine Vatto, Bas ter Weel, Isolde Woittiez and numerous seminar and conference participants. Remaining errors are our own. Finally, we are grateful to Statistics Netherlands for constructing the data set, and to the Ministry of Social Affairs and Employment for co-financing the construction of the data set.

1 Introduction

The consequences of the Great Recession and the ageing of the population have put the sustainability of public finances at risk. Governments are looking for cost-effective ways to increase labour participation. To this end, policymakers need good empirical information on how financial incentives affect the labour participation by different groups on the labour market, and the associated changes in public expenditures and revenues. Policymakers receive information from two different sources when considering the effectiveness of tax-benefit reforms. On the one hand, quasi-experimental studies report on the effects of specific reforms ('natural experiments') implemented in the past. Strong points of this approach are that it relies on minimal assumptions and has a transparent source of identification (Angrist and Pischke, 2009; Heckman, 2010). However, weak points are the absence of an underlying economic model, the external validity of the treatment effects and the limited scope for doing counterfactual policy analysis.[1] On the other hand, they get simulation results from analyses with structural models. The strong points of this approach are the use of an explicit economic model, the possibility to predict treatment effects in external environments and the wide scope for conducting counterfactual policy analysis (Heckman, 2010; Keane, 2010). But weak points of structural models are that they rely on more assumptions and that the source of identification is not always transparent. Looking at the strengths and weaknesses of both approaches, they can be said to complement each other.

In this study we combine structural models with quasi-experimental studies to evaluate the effectiveness of tax-benefit reforms in the Netherlands. Specifically, we estimate structural models for labour supply of a large number of subgroups on the Dutch labour market, compare the simulated treatment effects of policy reforms in the structural models with the results from three quasi-experimental studies, and then use the structural model to

[1] Some of these concerns are overcome in the so-called sufficient statistics literature (e.g. Chetty, 2009). In the sufficient statistics literature, authors use an explicit economic model to derive e.g. elasticities that are estimated in the programme evaluation literature. However, this approach can be used only for the analysis of counterfactual small reforms, and cannot be used for the counterfactual analysis of major reforms such as e.g. the introduction of a flat tax system.

study counterfactual policy reforms that feature prominently in the political debate in the Netherlands and abroad.

We estimate a structural discrete-choice model for labour supply, building on a large body of literature, e.g. Aaberge et al. (1999, 1995); Bargain et al. (2014); Brewer et al. (2006a); Keane and Moffitt (1998); Van Soest (1995). Discrete-choice models have the advantage of being able to take into account all the complexities in the budget set that result from the tax-benefit system (such as kinks and non-convexities). Furthermore, the estimation of discrete-choice models does not require ex-ante imposition of quasi-concavity of preferences, as this can be checked ex post.[2] We use an exceptionally large and rich administrative household dataset which was constructed specifically for this project by Statistics Netherlands, the Labour Market Panel (*Arbeidsmarktpanel*). Our sample consists of more than 840,000 observations and the size of this data set allows us to precisely estimate preferences over income, leisure and formal childcare (for parents with young children) for a large number of subgroups.

Next, we use the estimated structural model to simulate a number of key reforms implemented in the past and compare the simulated treatment effects with quasi-experimental studies on the same reforms. In particular, we compare the simulated treatment effects of the 2005–2009 reform of childcare subsidies and in-work benefits for households with young children with the estimated treatment effects presented in Bettendorf et al. (2015). Furthermore, we compare the simulated treatment effects of the 2002 reform of the in-work benefit for single parents with the estimated treatment effects presented in Bettendorf et al. (2014). Finally, we compare the simulated intensive margin (hours worked per employed person) elasticities of the structural model with the estimated intensive margin elasticities presented in Bosch and van der Klaauw (2012) and Bosch and Jongen (2013), who use the 2001 tax reform that substantially reduced marginal tax rates. Part of the variation used

[2]Studies using continuous labour supply choices and piecewise-linear budget constraints need to impose global quasi-concavity of preferences ex ante, which may have led to upward biased estimates of labour supply elasticities in these studies (MaCurdy et al., 1990).

in the quasi-experimental studies is also used in the estimation of the structural model, but part of the variation also comes from outside the data period used to estimate the structural model.[3]

Our main findings are the following. First, we uncover large differences in labour supply elasticities between demographic groups and decision margins. Indeed, there are large differences in labour supply elasticities by age of the youngest child. We also find large differences in the relative importance of the extensive margin (participation) and intensive margin (hours per employed person), with the intensive margin response still substantial for women with young children but small for most other groups. Second, we find that the structural model gives a good prediction of the treatment effects estimated in quasi-experimental studies on past reforms. Therefore, we feel confident to simulate counterfactual tax-benefit reforms using the structural model. Third, we find that reducing marginal tax rates is not an effective way to promote labour supply. In-work benefits targeted at low-wage earners appear more effective, and policies targeted at working mothers with young children generate the largest labour supply response. However, policies targeted at working mothers with young children generate little additional revenue for the government, because they already receive substantial subsidies when working in the baseline. Using the structural model we also simulate some large counterfactual tax reforms that feature prominently in the political debate. We find that proposals for a move towards a flat tax system, a basic income, or joint taxation are not effective. Indeed, an efficient tax system accounts for the large heterogeneity in behavioural responses we uncover, and hence cannot be too simple.

This study contributes to the existing literature in a number of ways. The size of our dataset allows us to estimate preferences separately for subgroups that earlier studies did not consider (previous studies typically only focussed on couples, e.g. Van Soest and Das, 2001; Van Soest et al., 2002) or had to be pooled due to the limited number of observations (previous studies typically pooled households with and without children, e.g. Bargain et al.,

[3]Hence, our comparison of the structural model with the quasi-experimental studies goes beyond a test of 'goodness-of-fit'.

2014). We consider a large number of household types, do not pool the data over the various household types, and uncover much more heterogeneity in the behavioural responses than previous studies. We also exploit detailed administrative data on the use and price of formal childcare, information that is not readily available in most other labour supply studies (e.g. Bargain et al., 2014; Blundell et al., 2000; Blundell and Shephard, 2012; Van Soest and Das, 2001). Furthermore, previous studies had to rely on one cross-section of data or on a few repeated cross-sections from a period when there was hardly any change in the tax system (e.g. Bargain et al., 2014; Van Soest and Das, 2001; Van Soest et al., 2002). Hence, identification in those studies comes only from cross-sectional differences in financial incentives due to non-linearities in the tax-benefit system. In contrast, we use several years of data, which includes a major reform of childcare subsidies and in-work benefits for working parents. These reforms strengthen our identification by generating large exogenous variation in the budget constraints. Also, we compare the predictions of the structural model with results from a number of quasi-experimental studies on key past reforms and thus contribute to the growing body of literature that evaluates the performance of structural models by comparing simulated policy responses with the results from (quasi-)experimental studies (e.g. Attanasio et al., 2011; Bargain and Doorley, 2017; Brewer et al., 2006b; Cai et al., 2008; Geyer et al., 2015; Hansen and Liu, 2015; Pronzato, 2018; Thoresen and Vatto, 2015; Todd and Wolpin, 2006). In addition, because our structural model is fully integrated with a detailed tax-benefit calculator, we are able to study the effectiveness of various reform proposals, taking into account the budgetary effects of the behavioural responses to the reform. Indeed, the integrated model allows us to go beyond back-of-the-envelope calculations on the effectiveness of different reform proposals (as in e.g. Blau, 2003; Lokshin, 2004).

The outline of the paper is as follows. Section 2 describes the Dutch labour market and policy environment. Section 3 outlines the structural model and empirical strategy. Section 4 describes the dataset used to estimate the structural model and Section 5 presents the

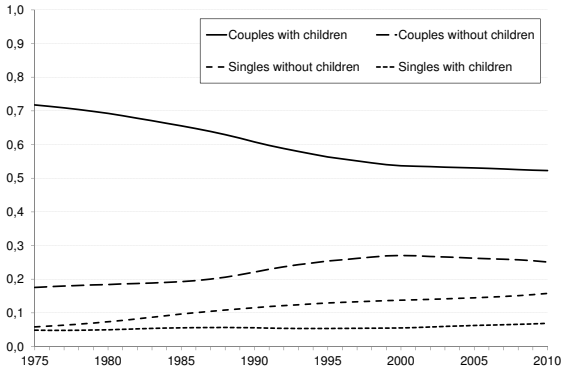
estimation results and corresponding labour supply elasticities. Section 6 then compares the simulated treatment effects of the structural model on a number of key past reforms with the estimated treatment effects from quasi-experimental studies. Next, Section 7 studies the effectiveness of a number of counterfactual tax-benefit reforms. Section 8 discusses our findings and concludes. Supplementary material is given in an appendix.

2 The Dutch labour market and tax-benefit system

Over the past decades, the Netherlands, like many other developed countries, witnessed a substantial shift in the household composition of the population, see Figure 1. In particular, the share of couples with children has declined, whereas the share of couples without children has increased. Furthermore, the share of singles, both with and without children, has increased. Hence, empirical knowledge of the behavioural responses of singles and single parents is becoming increasingly relevant.

Another important development has been the rise in the employment rate of women. Figure 2 shows the employment rate for men and women in the Netherlands over time, along with the development in the employment rate for a number of other countries. The employment rate for men in the Netherlands has increased somewhat over the 1975–2014 period, but with a noticeable dip in the 1980s due to the recession and the increased use of early retirement and disability benefits. Since the 1980s, early retirement benefits have been cut down substantially and have become more actuarially fair, and access to disability schemes has become more difficult. As a result, the employment rate of men has rebounded to pre-1980 levels. In 2014, the employment rate of men in the Netherlands was one of the highest of all OECD countries. The rise in the employment rate of women has been nothing short of spectacular, from just 30% in 1975 to 70% at the start of the Great Recession. Indeed, by 2014, the Netherlands had one of the highest employment rates for women in the OECD. An age-period-cohort analysis by Euwals et al. (2011) indicates that cohort

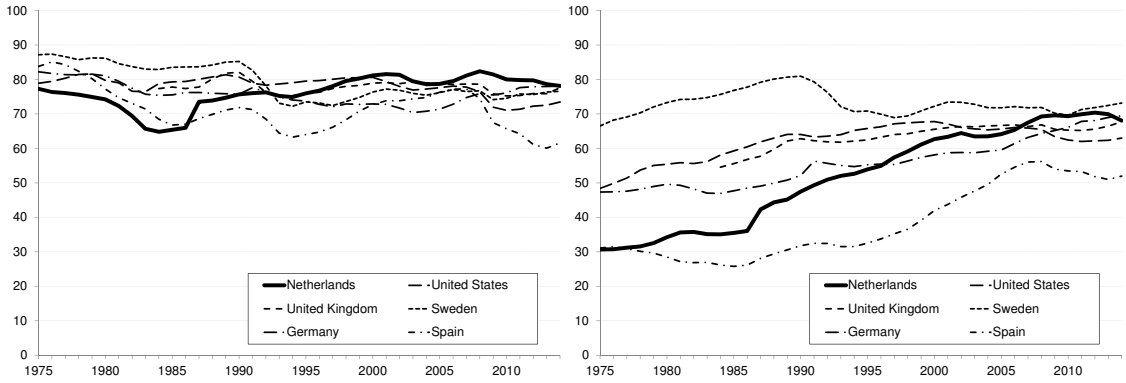
Figure 1: Shares of individuals per household type in the Netherlands



Source: Statistics Netherlands (statline.cbs.nl). Note: Individuals 15-64 years of age.

Figure 2: Employment rates in selected countries

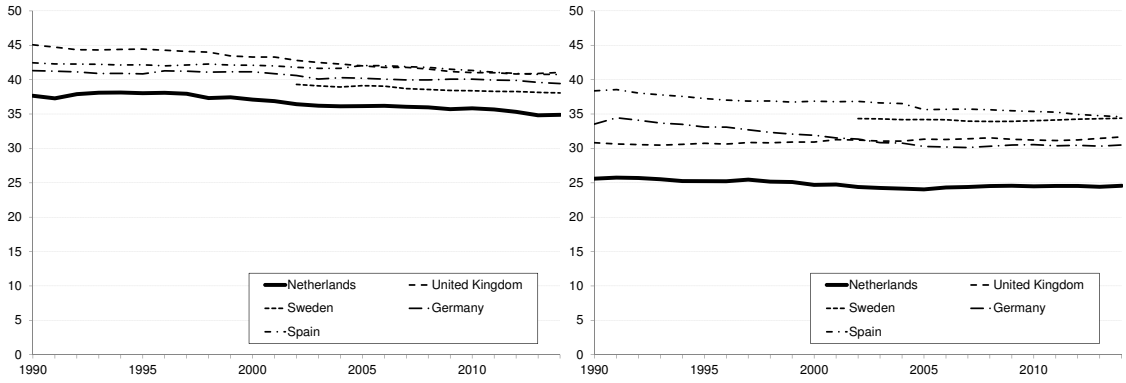
(a) Men (b) Women



Source: OECD (2016). Note: Individuals 15-64 years of age.

Figure 3: Average usual weekly hours worked on the main job in selected countries

(a) Men (b) Women



Source: OECD (2016). Note: Individuals 15-64 years of age.

effects are an important driver behind this increase, in particular for women with young children. The steep rise in the employment rate of women has important implications for the behavioural responses to changes in financial incentives. We estimate much lower labour supply elasticities for women in couples than previous studies for the Netherlands (e.g. Van Soest, 1995; Van Soest and Das, 2001) that used data of the 20th century.[4]

Whereas the rise in the employment rate of women in the Netherlands has been spectacular, their hours worked remained remarkably stable over the past decades, see Figure 3. If anything, hours worked per week by employed women decreased slightly. However, a noticeable gap of 5 to 10 hours per week remains with their peers in other European countries. The changes in hours worked per week by employed men were also limited over the past decades, with a slightly downward trend. However, also for men, the Dutch, on average, work a few hours per week less than their peers in other European countries. Below we study how changes in tax-benefit policies affect both the participation and the hours-worked-per-week decisions.

Considering the Dutch tax-benefit system, the Netherlands has an individual tax system, with specific tax credits and subsidies targeted at certain groups. However, income-dependent income support is based on household rather than individual income. The financial incentives implicit in the tax-benefit system are illustrated in the appendix, Section A, where we present so-called effective marginal tax rates (EMTRs) and participation tax rates (PTRs) for different subgroups.[5] The analysis shows that EMTRs and PTRs are particularly high for singles and single parents, and particularly low for secondary earners. Below we consider how changes in the EMTRs and PTRs following tax-benefit reforms affect the participation and hours-worked-per-week decisions of different groups on the labour market.

[4]This is in line with the findings for the US in Blau and Kahn (2007) and Heim (2007), they find that the female labour supply elasticity has declined substantially over the past decades, along with the rise in female employment rates. Furthermore, using a cross-section of countries, Bargain et al. (2014) show that the female labour supply elasticity is lower for countries that have a higher female employment rate.

[5]Following e.g. Brewer et al. (2010) and OECD (2016).

3 Structural model

We develop a structural model, where households are assumed to maximise a unitary household utility function. The most elaborate specification, for couples with young children, is outlined below. In this household, both partners choose their labour supply as well as their hours of formal childcare. The utility functions for the other household types (defined below) are a special case of this utility function.

The systematic part of household utility, U^s, depends on disposable income y, hours worked by the male h_m, hours worked by the female h_f, and hours of formal childcare cc. We ignore saving and borrowing, and hence consumption equals disposable income. For the functional form of U^s we use the translog specification:

$$\begin{aligned} U^s(\nu) &= \nu' \mathbf{A} \nu + \mathbf{b}' \nu + \mathbf{d}' \mathbf{1}[\mu > \mathbf{0}], \\ & \nu = (\log(y), \log(1 - h_m/T), \log(1 - h_f/T), \log(cc)), \\ & \mu = (h_m, h_f, c), \end{aligned} \tag{1}$$

with $\mathbf{A}$ being a symmetric matrix of quadratic coefficients and $\mathbf{b}$ being a vector of linear coefficients corresponding to the vector of the variables contained in ν. Note that we allow childcare to enter directly in the utility function (next to disposable income). The hours worked variables h_m and h_f in the vector ν have been transformed into indicators of leisure utilisation, representing the fraction of weekly time endowment T which is spent on activities unrelated to work. The vector $\mathbf{d}$ captures fixed costs of work, for the male and the female separately, and fixed costs of using formal childcare, for the household as a whole.

We allow for preference variation through observed individual and household characteristics $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ in parameters b_2, b_3 and b_4:

$$\begin{aligned} & \mathbf{b} = (b_1, b_2, b_3, b_4), \\ & b_1 = \beta_1, \quad b_2 = \mathbf{x}'_2 \beta_2 + \psi_2, \quad b_3 = \mathbf{x}'_3 \beta_3 + \psi_3, \quad b_4 = \mathbf{x}'_4 \beta_4 + \psi_4, \end{aligned} \tag{2}$$

which are the linear utility terms in leisure of the male, leisure of the female, and hours of formal childcare, respectively. The same variation is also allowed for the fixed costs parameters $\mathbf{d}$. We further allow for unobserved preference heterogeneity in the preference parameters for leisure (ψ_2 and ψ_3, for the male and female, respectively) and formal childcare (ψ_4).[6] We do not allow for observed and unobserved preference heterogeneity in the coefficient b_1 of income, because it is hard to identify this preference heterogeneity separate from the preference heterogeneity in leisure and childcare.

Disposable household income is given by:

$$y = w_m h_m + w_f h_f - T(w_m, h_m, w_f, h_f; q) - TC(p_{cc}, cc; q) + S(p_{cc}, cc, y_t; q), \tag{3}$$

where w_m and w_f denote the gross hourly wage for the male and the female,[7] $T(.)$ denotes taxes and employees' premiums, q denotes individual and household characteristics, $TC(.)$ is the total cost of formal childcare, with p_{cc} denoting its price per hour, and $S(.)$ is the childcare subsidy, which depends on the hourly price of formal childcare, the hours of formal childcare, taxable income y_t and household characteristics (like the ages of the children).

For workers, we observe gross hourly wages which are used to compute the work-related part of income for each alternative in the choice set.[8] For non-workers, we simulate wages using estimates from a model that accounts for selection (Heckman, 1979)[9], and taking multiple draws from the estimated wage error distribution, see Section B in the appendix. Similarly, for households that use formal childcare we use the observed hourly prices of formal childcare, and for non-users we simulate hourly prices using estimates from a model that accounts for selection and taking multiple draws from the estimated gross hourly price error distribution, see Section C in the appendix.

For our empirical specification we use a discrete-choice model. Households choose their

[6] We use Halton sequences to draw the random terms (Train, 2003). For simplicity, we assume that there is no correlation between these unobserved preference heterogeneity terms.

[7] For simplicity we assume that the gross hourly wage does not depend on the hours worked.

[8] We use administrative data on hours worked and wages, hence measurement error is less of a concern.

[9] Here we follow e.g. Blundell et al. (2007) and Bargain et al. (2014).

preferred combination of hours of work and hours of formal childcare from a finite set of alternatives $j \in \{1, ..., J\}$. Next to the systematic part $U^s(\nu_j)$, the utility function contains alternative-specific stochastic terms ε_j:

$$U(\nu_j) = U^s(\nu_j) + \varepsilon_j. \tag{4}$$

These stochastic terms are assumed to be independent and identically distributed across alternatives, and to be drawn from a Type 1 Extreme Value distribution. This leads to a multinomial logit specification (McFadden, 1978).

Random preference heterogeneity, along with the draws from the estimated wage and price equations for non-workers and non-users of formal childcare, respectively, complicate the estimation of the likelihood. We use $R = 50$ (independent) draws from the wage distribution and for non-working men and women, the price distribution for non-users of formal childcare and the random terms for unobserved heterogeneity.[10] We use simulated maximum likelihood, where the likelihood is given by:

$$L = \prod_{i=1}^{N} \frac{1}{R} \sum_{r=1}^{R} \left(\frac{\exp(U_k^i(w_{i,m,r}, w_{i,f,r}, p_{cc,r}, \psi_{i,2,r}, \psi_{i,3,r}, \psi_{i,4,r}))}{\sum_{j=1}^{J} \exp(U_j^i(w_{i,m,r}, w_{i,f,r}, p_{i,cc,r}, \psi_{i,2,r}, \psi_{i,3,r}, \psi_{i,4,r}))} \right)^{D_{ki}}, \tag{5}$$

with D_{ki} being an indicator function taking the value 1 for the observed choice for household i, and zero otherwise.[11]

[10] The number of draws is kept relatively low to limit the computational complexity of the model. However, increasing the number of draws did not change the predictions of the structural model.

[11] Note that for workers and users of formal childcare we take the actual gross hourly wage and actual hourly price, respectively, for each draw r.

4 Data structural model

We use data from the Labour Market Panel (LMP) of Statistics Netherlands (2012). This dataset was constructed specifically for the empirical analysis presented here. The LMP is a large administrative household panel dataset over the period 1999–2009, containing a rich set of individual and household characteristics, including gender, month and year of birth, the level of education and ethnicity of all adult household members, the ages of the children and place of residence. In addition, the LMP also contains administrative data on gross income from different sources (e.g. wages, profits, benefits) and on hours worked. Finally, the LMP contains administrative data on the use and gross hourly price of formal childcare for each child in formal childcare.[12] Because data on childcare is only available from 2006 onwards, we restrict the sample to the 2006–2009 period.

In the empirical analysis, we model the labour supply decision for employed people, those on welfare benefits, and those without personal income.[13] We make a number of additional selections. We exclude people under 18 years of age (most of them are in education), and those over 63 years of age (we do not model the retirement decision). Furthermore, we do not model the labour supply decision of students, people on retirement or disability benefits and the self-employed. Below these people are referred to as having an 'inflexible' labour supply. They are not included, because reliable information is not available on their hours worked or because we are unable to determine their budget constraint. Furthermore, same sex households are also excluded, as are households for which characteristics about individual members or the household are missing.[14] In the end, we use 840,348 observations in the

[12]Unfortunately, informal childcare is not included in our administrative dataset. However, De Boer et al. (2015) estimate preferences using the overlap in working hours of parents minus the hours of formal childcare as a proxy for informal childcare. The resulting labour supply and formal childcare price elasticities are very similar to the model without the proxy for informal childcare.

[13]We remove people on unemployment benefits from the structural model, implicitly assuming that they are constrained in their labour supply choice. In the simulation model, we add a reduced-form model on the probability of people being on unemployment benefits, following Ericson and Flood (2012), see Section H in the appendix.

[14]For couples with a youngest child of between 0 and 3 and between 4 and 11 we use a 50% subsample. These groups have the largest discrete-choice set, and using the full sample was not possible due to memory restrictions in Stata.

estimations.

We estimate structural discrete-choice models for the following 15 household types: childless singles, single parents with a youngest child aged 0–3, 4–11, 12–17 or 18 years of age or older, respectively, adult children living with their parent(s)[15], couples without children with both partners flexible, couples without children where only the man is able to adjust his labour supply (the woman is 'inflexible', as defined above), couples without children where only the woman is able to adjust her labour supply (the man is 'inflexible'), couples where both partners can adjust their labour supply (are 'flexible') and with a youngest child aged 0–3, 4–11, 12–17 or 18 years of age or older, respectively, couples with children where only the man can adjust his labour supply (the woman is 'inflexible'), and couples with children where only the woman can adjust her labour supply (the man is 'inflexible').

We discretise the data for the discrete-choice model in the following way. Adults can choose from 6 labour supply options: working 0, 1, 2, 3, 4 or 5 days per week, each day equalling 8 hours.[16] For childcare, we allow for 0, 1, 2 and 3 days[17], with data showing a typical childcare day to equal 10 hours.[18], and a typical out-of-school-care day equals 5 hours.[19] Couples with a youngest child aged 0 to 3 or 4 to 11 have the largest choice set: $6 \cdot 6 \cdot 4 = 144$ alternatives.

To determine disposable household income in each discrete option we use the advanced tax-benefit calculator MIMOSI (Koot et al., 2016).[20] MIMOSI is the official tax-benefit calculator of the Dutch Government for the (non-behavioural) analysis of the impact of reform proposals on the disposable income distribution and the government budget. MIMOSI takes into account all (national[21]) taxes, social security premiums, and income independent

[15]We model adult children living with their parents as a separate household category that is not entitled to welfare benefits, in accordance with Dutch legislation.

[16]Classified as: $0 \in [0, 5>, 8 \in [5, 13>, 16 \in [13, 21>, 24 \in [21, 29>, 32 \in [29, 37>, 40 \in [37, \infty>$.

[17]The data show that using formal childcare for more than 3 days per week is rare in the Netherlands. The remaining childcare needs are usually met by informal care or parents themselves.

[18]Classified as: $0 \in [0, 0], 10 \in <0, 15>, 20 \in [15, 25>, 30 \in [25, \infty>$.

[19]Classified as: $0 \in [0, 0], 5 \in <0, 7.5>, 10 \in [7.5, 12.5>, 15 \in [12.5, \infty>$.

[20]Disposable incomes in the estimations and simulations are in 2006 prices. We use the CPI to convert nomimal values in later years into 2006 prices.

[21]Local taxes account for only a small portion of total taxes in the Netherlands (3.3% in 2007, European

subsidies and tax credits. Furthermore, MIMOSI also calculates the childcare subsidy in each option.[22] Furthermore, in accordance with the law, we ensure that household disposable income (excluding childcare costs and childcare subsidies) cannot drop below the welfare level. For each discrete option we also calculate the net transfer from the household to the government (positive or negative). This allows for an accurate calculation of the net budgetary costs of the reforms we simulate.

5 Estimation results structural model

For each household type, we allow preferences for leisure to depend on age and fixed costs of work to depend on the level of education (in three classes) and ethnicity (in three classes). Furthermore, for households with a youngest child aged 0–3 or 4–11, we allow preferences for the use of formal childcare to depend on the level of education and ethnicity, and whether or not their residential location was in a large city (>150,000 inhabitants). The preference parameters are not discussed here, because it is often a combination of preference parameters that drives behavioural responses (the estimated preferences are given in Section E in the appendix).[23] Furthermore, there is no analytical solution for the labour supply elasticity in discrete-choice models. Therefore, following the literature (e.g. Bargain et al., 2014), we simulate these elasticities by increasing gross wages by 10%. We present the total elasticity (% change in total hours worked over the % change in the gross wage rate), and the decomposition into the extensive margin elasticity (% change in the participation rate over the % change in the gross wage rate) and the intensive margin elasticity (% change in hours worked per employed person over the % change in the gross wage rate).

Figure 4 gives the simulated labour supply elasticities for couples where both partners can

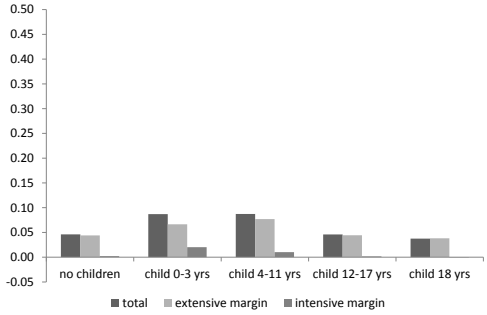
Union, 2014).

[22]Only working parents are entitled to receive the childcare subsidy, with the subsidy level depending on the gross hourly price of childcare per type of childcare (daycare or out-of-school care), and only up to a maximum price beyond which parents receive no additional subsidy, household income (subsidies are lower for higher incomes), and number of children (subsidies are higher for second, third and subsequent children in formal childcare).

[23]The models generate a good fit of the hours distribution, see Section E in the appendix.

Figure 4: Households with two 'flexible' persons

(a) Men



(b) Women

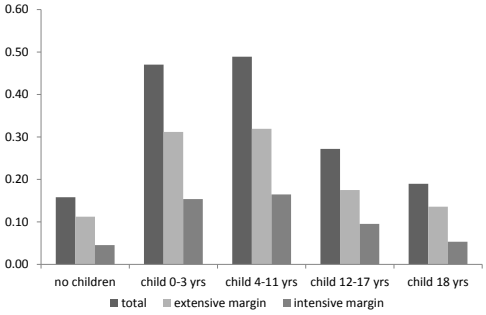
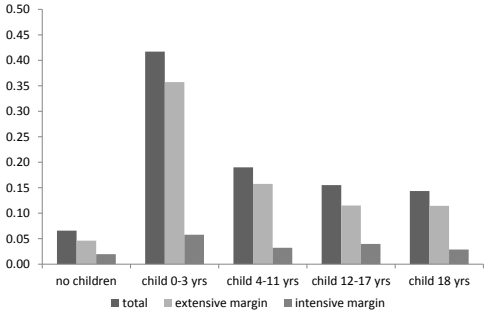
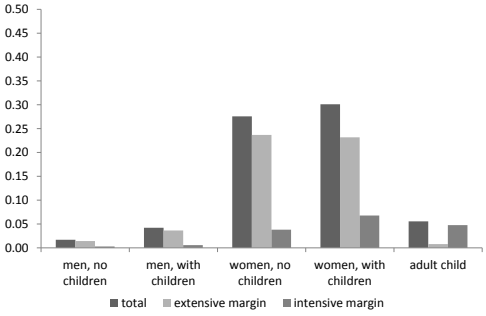


Figure 5: Households with one 'flexible' person, and adult children

(a) Singles and single parents



(b) Individuals with an inflexible partner, and adult children living at home



adjust their labour supply (are 'flexible'). We find small, positive labour suppy elasticities for men. Labour supply elasticities are higher for women, on the extensive margin but also on the intensive margin. Furthermore, labour supply elasticities for women in couples are particularly high when they have a young child.[24]

Figure 5 shows that the labour supply elasticity is relatively low for childless singles without children.[25] The labour supply elasticity is much higher for single parents with a

[24]Figure G.1 in the appendix gives the cross-elasticities, e.g. the % change in total hours worked by one partner over the % change in the gross wage rate of the other partner. Cross-elasticities are close to zero for men, but are negative and non-negligible for women.

[25]We also estimated preferences separately for single men and single women, but they were almost identical,

youngest child of 0–3 years of age. The labour supply elasticity of single parents whose youngest child is over the age of 3 is lower, though still higher than that of childless singles. Also note that the differences among single parents are primarily driven by differences in the extensive margin elasticity. The intensive margin response for single parents is quite small.[26]

Figure 5 also gives the labour supply elasticities for couples with just one 'flexible' partner. For these groups, we pool couples with children of all ages. Most men with an 'inflexible' partner are working, and typically also full-time (see Section E in the appendix). Hence, there is little upward potential in terms of total hours worked, and they have a relatively low labour supply elasticity. For women there is more upward potential in total hours worked. Women with an 'inflexible partner' have a higher labour supply elasticity. Finally, adult children living with their parents generally have a high participation rate, resulting in a relatively low labour supply elasticity.[27]

so we pooled these groups in the estimations.

[26]Their budget constraint plays an important role here, as working extra hours or days may generate little additional net income due to their relatively high effective marginal tax rate, see Section A in the Supplementary material.

[27]Comprehensive surveys of the labour supply elasticity in a large number of countries can be found in Blundell and MaCurdy (1999), Bargain and Peichl (2017) and the recent estimates presented in Bargain et al. (2014). We compare our results with the recent estimates for Europe and the United States in Bargain et al. (2014). Bargain et al. (2014) find that for women in couples the total hours elasticity ranges from 0.1 to 0.6 among countries (with a mean of 0.27). Our estimates for women in couples with young children fall within this range. The estimates for women in couples with older children or no children are somewhat lower. However, the participation rate of women in the Netherlands is relatively high from an international point of view. For men in couples, the total hours elasticity in Bargain et al. (2014) ranges from 0.05 to 0.15 among countries (with a mean of 0.10). Our estimates for men in couples without and with children are on the lower end of that range. For single men, Bargain et al. (2014) find a total hours elasticity ranging from of 0.0 to 0.4 (and some even higher). For single women they find an elasticity ranging from 0.1 to 0.5 (and again some even higher). Our estimates for singles are on the lower end of that range, and our results for single parents are more in the middle and upper part. Bargain et al. (2014) find that the extensive margin elasticity is typically (much) more important than the intensive margin elasticity and that cross-elasticities for women in couples are non-negligible, and are close to zero for men in couples. These are also our findings.

6 Comparison predictions structural model with results from quasi-experimental studies

Next, we consider whether the structural model can predict the effects of past reforms. There is a growing body of literature that compares the simulated policy responses in structural models with the results from (quasi-)experimental studies (e.g. Attanasio et al., 2011; Bargain and Doorley, 2017; Brewer et al., 2006b; Cai et al., 2008; Geyer et al., 2015; Hansen and Liu, 2015; Pronzato, 2018; Thoresen and Vatto, 2015; Todd and Wolpin, 2006). We present three such comparisons: i) one for couples with a youngest child of 0–3 and 4–11 years of age, ii) one for single parents with a youngest child of 12–15 years of age, and iii) one for the intensive margin responses for a number of household types.

Table 1 shows our structural model results for couples with a youngest child of 0–3 and 4-11 years of age. Bettendorf et al. (2015) use differences-in-differences (DD) to analyse the employment effects of a combination of reforms during the period 2005–2009 targeted at households with children of 0–11 years of age. These reforms contained three elements: 1) an increase in childcare subsidies (column *Childcare subsidies*), 2) an increase in the in-work benefit for secondary earners with children of 0–11 years of age (column *Income-depend.combi.credit*), and 3) the in-work benefit for both primary and secondary earners with children of 0–11 years of age was abolished (column *Combi.credit*).[28] Bettendorf et al. (2015) use data for the 1995–2009 period from the Labour Force Survey. They present estimation results for women in couples with a youngest child of 0–11 years of age. To make the comparison with the DD as clean as possible, we use the same sample as Bettendorf et al. (2015) to estimate the treatment effects by the subgroups with a youngest child of 0–3 or 4–11 years of age, respectively. The results are given in Table 1, along with the simulation results from the structural model. The results for the structural model are consistent with the results from the DD analysis for women with children. The estimated effects on the

[28] See Bettendorf et al. (2015) for a detailed description of the reforms.

Table 1: Quasi-experimental check: couples with young children

	Structural model				DD analysis[a]	
	Childcare subsidies	Combi. credit	Income-depend. combi. credit	**Total**	**Coefficient**	SE
			Absolute changes			
Youngest child 0-3 yrs						
Participation rate women	0.019	–0.006	0.020	**0.033**	**0.020**	0.007
Hours worked per week women	0.752	–0.106	0.630	**1.275**	**1.099**	0.215
Participation rate men	0.003	–0.002	0.004	**0.005**	**0.006**	0.004
Hours worked per week men	0.067	–0.033	0.046	**0.080**	**–0.364**	0.229
Youngest child 4-11 yrs						
Participation rate women	0.003	–0.005	0.016	**0.014**	**0.022**	0.007
Hours worked per week women	0.105	–0.086	0.453	**0.471**	**0.741**	0.212
Participation rate men	0.000	–0.001	0.003	**0.002**	**0.003**	0.004
Hours worked per week men	0.018	–0.024	0.051	**0.045**	**–0.141**	0.227

[a] Additional estimates using the same sample as Bettendorf et al. (2015), full regression results available on request.

Table 2: Quasi-experimental check: single parents

	Structural model	DD analysis[a]		RD analysis[a]	
		Coefficient	SE	Coefficient	SE
		Absolute changes			
Participation rate					
Youngest child of 12–15 years	**0.000**	**0.004**	0.006	**0.004**	0.011
Youngest child of 4–11 years	–0.006				
Youngest child of 0–3 years	–0.011				

[a] For details of the DD and RD results, see Bettendorf et al. (2014).

participation rate of men with children is also in line with the prediction from the structural model. The only coefficient of the DD analysis which differs somewhat from the prediction of the structural model is the intensive margin response by men, for which the DD analysis suggests a larger, negative though not statistically significant response than the structural model.

Table 2 shows the evaluation of the predictions made by our structural model for single parents with a youngest child of 12–15 years of age. Bettendorf et al. (2014) use DD and regression discontinuity (RD) to study the impact of an in-work benefit targeted at single parents. In 2002, the children's age of eligibility was increased, and the target group of the in-work benefit was extended from single parents with a youngest child of 0–11 years of age to single parents with a youngest child of 0–15 years of age. The outcome of the analysis is that they find a small effect on labour participation, in both the DD and RD analyses, not significantly different from zero. Table 2 shows the effect of abolishing the in-work benefit targeted at single parents. We simulate the effect on the participation rate of single parents with a youngest child of 0–3, 4–11 and 12–15 years of age.[29] We find the effect on the participation rate with –1.1 percentage points to be quite sizeable for single parents with a youngest child of 0–3 years of age, whereas it drops to –0.6 percentage points for single parents with a youngest child of 4–11 years of age, and to 0.0 percentage points for single parents with a youngest child of 12–15 years of age. Indeed, single parents with an older youngest child are already less responsive to financial incentives than single parents with a younger youngest child. More importantly, again, the structural model predicts treatment effects that are in line with the results from the quasi-experimental study.

Finally, we compare the intensive-margin responses in the structural model with a quasi-experimental study on intensive-margin responses. Figure 4 and 5 show that intensive-margin responses are typically rather small, and much smaller than extensive-margin responses. We compare the intensive-margin elasticities with results from the DD analysis in Bosch

[29]Bettendorf et al. (2014) do not consider the effect of the reform on hours worked.

and Jongen (2013). They use the 2001 tax reform, which generated large heterogeneous variation in marginal tax rates. For men in couples, they find very low intensive-margin elasticities with a point estimate of 0.00 (s.e. 0.01), in line with the structural model. For women in couples, they find larger intensive-margin elasticities, with a point estimate of 0.15 (s.e. 0.06).[30] This is in line with the structural-model results on the response at the intensive margin for women with young children, and somewhat higher than for the other groups of women in couples. For singles and single parents, Bosch and Jongen (2013) also find somewhat higher intensive-margin elasticities than the structural model, 0.15 to 0.20, respectively. We should note though that the comparison is somewhat complicated because we compare gross wage elasticities of the structural model [31] to the net wage elasticities of the DD. The latter are typically higher (Bargain et al., 2014).

7 Simulating tax-benefit reforms

Overall, the structural model predicts the treatment effects of past reforms rather well. We then exploit the strength of the structural model by simulating counterfactual tax-benefit reforms. We first consider changes in a selected set of policy parameters, to illustrate which policies are more or less cost-effective in terms of stimulating labour supply. Subsequently, we consider a number of major tax reforms that feature prominently in the policy debate in the Netherlands and abroad.[32]

[30] Bosch and van der Klaauw (2012) apply the estimator of Blundell et al. (1998) to estimate the intensive-margin responses for women in couples, using data from the Labour Force Survey. They find a coefficient insignificantly different from zero.

[31] Which is common in the literature on discrete-choice labour supply, see for instance Bargain et al. (2014).

[32] Before we can simulate these reforms we have to prepare the model for policy simulation. We estimated the preferences of the different household types using data from the LMP. However, CPB uses data from the Income Panel of Statistics Netherlands to calculate the effects of tax-benefit reforms on the income distribution and the government budget (see Koot et al., 2016). To have a single model that will generate all relevant output, we integrate the discrete-choice model for labour supply and childcare with the tax-benefit calculator using the Income Panel data. Section H in the appendix describes this process.

7.1 Changes in selected tax-benefit policies

Marginal tax rates We first consider the effects of changes in marginal tax rates. More specifically, we consider the effects of decreasing the tax rate of the first, second, third and fourth (open) tax bracket, so that, in each simulation, tax revenues decrease by 1.5 billion euros before behavioural responses. Table 3 gives the results in columns (1)–(4), respectively.[33]

We report the effects on hours worked per week and on the participation rate. Hours worked per week includes the zeros for the non-employed. The participation rate is the number of persons employed over the total number of employed and not employed (excluding the 'inflexible' people). We also calculate the effect on labour productivity per hour worked, which is obtained by subtracting the change in hours worked from the change in labour costs, where the change in labour costs, in turn, is an approximation for the change in output. Furthermore, we also calculate the so-called knock-on effects for the government budget, i.e. the net budgetary savings due to behavioural responses, expressed as a percentage of the ex-ante (before behavioural responses) budgetary 'shock'. An alternative strategy would be to simulate policy reforms that are budgetary neutral after taking into account behavioural responses. However, this is rather time consuming, and does not change the relative effectiveness of the different policies. Finally, we report the effect on household income inequality (before behavioural changes), using the Gini coefficient.[34]

Column (1) gives the results for the decrease in the tax rate in the first tax bracket. Overall, hours worked and the participation rate hardly change. However, this is the net

[33]To keep the table to a manageable size, we aggregate the results to some broader categories. Specifically, we use the following aggregates: 1) 'Men in couples youngest child 0–17' and 'Women in couples youngest child 0–17' which are respectively men and women in 'two-flex households' with a youngest child of 0–17 years of age, 2) 'Men in other couples' and 'Women in other couples' are men and women in 'one-flex households' without children, where one of the partners is 'inflexible' (couples with a youngest child of 18 years of age or older are also classified as couples without (dependent) children), 3) 'Single parents youngest child 0–17' and 4) 'Singles' consists of singles without (dependent) children and adult children living with their parents (this group also includes single parents with a youngest child 18 years of age or older). Furthermore, the total results over all groups are for people whose labour supply is determined within the model only, so excluding the 'fixed' labour supply of the 'inflexible' partners.

[34]Where household income is weighted by the equivalence scales of Statistics Netherlands.

Table 3: Changes in selected tax-benefit policies

Simulation	Base	(1) First bracket[a]	(2) Second bracket[b]	(3) Third bracket[c]	(4) Fourth bracket[d]	(5) Welfare benefits[e]	(6) In-work tax credit[f]	(7) Combination credit[g]	(8) Childcare subsidies[h]	(9) Child benefit[i]
Ex-ante government spending (in billion euros)		1.5	1.5	1.5	1.5	–0.5	1.5	0.5	0.5	–0.5
	Levels	Changes in percentages								
Hours worked per week	26.8	0.00	0.09	0.08	0.03	0.64	0.19	0.17	0.19	0.28
– Men in couples y. child 0–17	35.0	-0.08	0.05	0.12	0.09	0.76	–0.04	0.04	0.07	0.51
– Women in couples y. child 0–17	17.0	0.15	0.20	-0.03	-0.11	0.82	0.76	1.32	1.71	1.13
– Men in other couples	33.9	-0.07	0.02	0.07	0.07	0.53	–0.05	0.00	0.00	0.06
– Women in other couples	19.4	0.05	0.17	0.11	0.02	0.80	0.37	0.06	0.00	0.19
– Single parents y. child 0–17	19.8	-0.04	0.21	0.14	0.05	2.02	0.36	0.69	0.28	0.55
– Singles	30.0	0.03	0.10	0.07	0.02	0.36	0.22	0.00	0.00	0.00
Participation rate	0.82	0.02	0.04	0.00	-0.02	0.63	0.25	0.17	0.16	0.26
– Men in couples y. child 0–17	0.93	-0.04	0.06	0.04	0.00	0.78	0.14	0.11	0.09	0.45
– Women in couples y. child 0–17	0.77	0.14	0.00	-0.17	-0.14	0.71	0.61	0.79	0.92	0.84
– Men in other couples	0.91	-0.04	0.02	0.04	0.03	0.51	0.04	0.00	0.00	0.05
– Women in other couples	0.70	0.07	0.07	0.01	-0.02	0.75	0.40	0.06	0.00	0.16
– Single parents y. child 0–17	0.68	-0.02	0.11	0.04	0.01	1.97	0.46	0.61	0.20	0.39
– Singles	0.86	0.01	0.05	0.04	0.01	0.33	0.17	0.00	0.00	0.00
Labour productivity per hour		-0.02	-0.01	0.03	0.06	–0.10	–0.11	–0.03	–0.18	–0.05
Knock-on effects[j]		-4	8	11	13	119	0	13	–41	52
Gini coefficient[k]	0.24	–0.22	0.30	0.87	1.33	0.78	–0.35	–0.01	0.01	0.43

[a]Decrease in bracket rate of –0.8%. [b]Decrease in bracket rate of –2.1%. [c]Decrease in bracket rate of –3.5%. [d]Decrease in bracket rate of –5.2%. [e]Reduction in welfare benefits of –14%. [f]An increase in the in-work tax credit of 440 euros, by increasing the phase-in rate from 19.7% to 23.9%. The higher in-work tax credit is phased out from annual incomes of 34,000 euros onwards at 4%. [g] Increase in the combination credit for working secondary earners and working single parents with a youngest child of up to 12 years of age by 1,109 euros, by increasing the phase-in rate from 4% to 8%. [h] An increase in childcare subsidies for working parents, the change in childcare subsidies is set in such a way that there is a proportional decline of 38% in the parental contribution rate. [i] A decrease in the income-dependent child benefit for parents with a youngest child of up to 18 years of age. The subsidy is phased out from 19.676 euros at a rate of 6.75%. We decrease the subsidy by 45%, and keep the phase-out rate the same. [j] Knock-on effect as a percentage of the initial additional government expenditures, for people in the labour supply sample. [k]Gini coefficient of disposable household income, using equivalence scales. The Gini coefficient is calculated over the full Dutch adult population with a gross income of over 66% of the annual minimum wage.

result of some groups that decrease and some groups that increase their labour supply. In particular, there is a modest decrease in hours worked by men in couples, due to an income effect (for them, the first tax bracket is typically inframarginal), and a modest increase in hours worked by women in couples, for them the substitution effect dominates the income effect for them. Income inequality, as measured by the Gini coefficient, decreases.

Column (2) gives the effect of lowering the tax rate in the second tax bracket. The effect on overall labour supply is positive but modest. Men in couples now also work a bit more due to the substitution effect, while the effect on women in couples is larger than under reform (1). When comparing the effects on total hours worked per week with those on the participation rate, most of the response comes from the intensive margin rather than the extensive margin. Income inequality increases somewhat, as the lowest incomes do not benefit from a lower second tax bracket rate.

Column (3) then considers the effects of a decrease in the third tax bracket rate. The increase in overall labour supply in hours is somewhat smaller than in reform (2), because of the smaller effect on hours worked by women. Indeed, although for some of these women the third tax bracket is the relevant marginal tax bracket, their own income effect and the income effect from a higher income of their partner dominates.[35] For single parents and singles, we find a positive effect on labour supply in hours, they do not have an income effect from a partner and the substitution effect of the lower marginal tax rate dominates. Income inequality increases more than under reform (2).

Finally, column (4) gives the effects of lowering the tax rate in the fourth (open) tax bracket. This has only a small positive effect on overall hours worked, and the effect on labour supply in persons is negative (due to the 'added worker effect'). But, where the increase in hours worked is much smaller under reform (4) than reform (3), labour productivity increases more due to a composition effect, workers with income in the fourth tax bracket are more productive. Also, because high-income individuals pay a relatively large amount of taxes,

[35]The cross-effect of a higher income of men in couples on the labour supply of women in couples is an illustration of the 'added worker effect' (Lundberg, 1985).

the knock-on effect for the government budget is higher than for reforms (1)–(3). Lowering the top rate leads to the biggest increase in inequality of reforms (1)–(4).

Participation tax rates Next, we consider policy reforms targeted more at the 'participation tax rate', the effective tax on the transition from non-employment to employment. Specifically, we consider lowering the participation tax rate through a 'carrot' or a 'stick'.

Column (5) gives the simulated effects of a reduction in welfare benefits by 14% (the stick), for a total amount of 500 million euros. This leads to a substantial increase in overall labour supply, both in total hours worked and in persons, of +0.6%. The effects are much larger than the reforms considered before, because welfare benefits operate on the extensive margin. The response is particularly large for single parents; 32% of single parents are on welfare benefits in the base. The knock-on effects for the government are very high, because there is a sizeable reduction in the expenditures on welfare benefits due to behavioural responses. On the downside, this simulation causes a steep rise in income inequality.

In column (6) we use the 'carrot' instead, and consider an increase in the (general) in-work tax credit for a total amount of 1.5 billion euros, targeting the increase at low-income workers.[36] This also has a larger effect on total hours worked than reducing tax bracket rates because it is targeted at the extensive margin. Also on the upside, income inequality decreases, as the reform targets low income workers. Indeed, in this simulation, there is an increase in hours worked as well as a decrease in income inequality (see also Saez, 2002, for the potential welfare gains from in-work tax credits for low-income workers). On the downside, there is a sizeable drop in average productivity, due to a change in the workforce composition, and the knock-on effect is also close to zero.

Subsidies for households with young children Finally, we consider reforms targeted at households with young children. This group is of particular interest because there are many policies targeted at this group, and because mothers with young children appear particularly

[36]Specifically, we increase the maximum in-work tax credit by 441 euros, and start the phase-out at a rate of –4% as in the base, but from an annual income of 34,000 euros as opposed to 49,770 euros in the base.

responsive to changes in financial incentives. In these simulations, we use a smaller increase in government expenditures than before, because these reforms target only a subgroup of the working age population.

In column (7) we increase the income-dependent part of the in-work tax credit for secondary earners and single parents with a youngest child of up to 12 years of age.[37] This leads to a substantial increase in the number of hours worked given the budgetary impulse, because it targets the groups with highest labour supply elasticity. However, the knock-on effects are still limited, as secondary earners and working single parents with a young child already receive large subsidies in the base.

In column (8) we increase childcare subsidies. In particular, we consider a proportional decrease of 38% in the parental contribution that results after deducting the subsidy from the full hourly price. Again, there is a substantial increase in hours worked. The effect on total hours worked is somewhat larger than for reform (7). However, the childcare reform also leads to substantial substitution of informal care for formal care. As a result, the knock-on effect for the government budget is negative, making this reform less cost-effective than reform (7).

Moving from carrots to sticks, in column (9), we decrease the income-dependent child benefit for parents with small children.[38] This also leads to a substantial increase in hours worked and in the participation rate.[39] However, again, the downside of this 'stick' reform is that it increases inequality, as we take benefits away from low-income households.

[37]'Inkomensafhankelijke Combinatiekorting' in Dutch, which depends on income, with a phase-in rate of 4% for income above the threshold of 4,857 euros, reaching its maximum of 1,119 euros at a personal income level of 32,832 euros. We increase the maximum credit level by 1,100 euros, via an increase in the phase-in rate of 4 percentage points.

[38]'Kindgebonden Budget' in Dutch, this benefit is a fixed amount, which however depends on the number and ages of the children, up to a gross annual household income of 19,676 euros, after which it is phased out at a rate of 6.75% in the base. We decrease the maximum amount for all families by 45% and keep the phase-out rate the same.

[39]Note that there is also a small effect on men and women in other couples, these are the men and women in couples with a partner whose labour supply is fixed, but who have a dependent child.

7.2 Major tax reforms

After considering changes in single policy instruments, we now consider a number of major tax reforms that feature prominently in the current policy debate, in the Netherlands and abroad. Specifically, we consider the introduction of a flat tax system, a basic income system, and a move towards joint rather than individual taxation.[40] The simulation results are given in Table 4.

In column (1), we change the four tax brackets rates in the baseline, 36.5%, 42%, 42% and 52%, respectively, to a flat tax rate of 39.7%. This scenario is budgetary neutral before behavioural changes. We see that this flat tax increases overall labour supply. Men in couples and singles without children increase their labour supply, because the flat tax rate is lower than their initial marginal tax bracket rate. For women in couples, there are two opposing effects. First, women paying taxes in the first tax bracket now face a higher marginal tax rate and some women withdraw from the labour force. Second, women paying taxes in the second and higher tax bracket now face a lower marginal tax rate, triggering a positive response at the intensive margin. For women in couples, with children aged 0–17, the first effect dominates, whereas for women in other couples the second effect dominates. Finally, single parents increase their labour supply, which is caused by the sharp drop in the net welfare benefit due to the increase in the first bracket rate. The downside of this flat tax proposal is a substantial increase in income inequality.

In column (2), we therefore consider a flat tax system that is budgetary neutral *and* 'Gini neutral'. Specifically, we introduce a lump-sum subsidy for all adults of 1.950 euros and finance this with a flat tax rate of 45.3%. We find that to arrive at the same income inequality as in the baseline, the flat tax reduces labour supply.[41] Indeed, the flat tax

[40] Recently, CPB used the model to evaluate the governments Tax Plan for 2016 (Ministry of Finance, 2015). The aim of the Tax Plan was to create more employment by reducing the tax burden on labor with 5 billion euros. Important policy measures in the tax plan concerned a higher tax credit for all workers, a higher tax credit for working parents and a higher child care subsidy. There has been much debate among politicians and in the media about the employment effects of the new Tax Plan, which came into effect on 1 January 2016. Our model predicts that labor supply increases with 35,000 full-time equivalents (FTEs).

[41] A point made earlier by Jacobs et al. (2009).

Table 4: Major tax reforms[a]

Simulation	Base	(1) Flat tax[b]	(2) Flat tax, Gini neutral[c]	(3) Basic income[d]	(4) Joint taxation[e]
Ex-ante government spending (in billion euros)		0.0	0.0	0.0	0.0
	Levels		Changes in percentages		
Hours worked per week	26.8	0.24	–0.88	–5.47	–2.18
– Men in couples youngest child 0–17	35.0	0.66	–0.26	–3.23	0.14
– Women in couples youngest child 0–17	17.0	-0.69	–4.69	–21.5	–16.3
– Men in other couples	33.9	0.47	–0.02	–2.53	0.00
– Women in other couples	19.4	0.10	–1.72	–8.51	–3.12
– Single parents youngest child 0–17	19.8	0.61	–0.08	–2.10	–0.74
– Singles	30.0	0.11	–0.38	–2.65	–0.31
Participation rate	0.82	–0.07	–0.60	–4.56	–2.72
– Men in couples youngest child 0–17	0.93	0.24	–1.44	–2.87	–0.46
– Women in couples youngest child 0–17	0.77	–1.04	–0.36	–16.3	–13.9
– Men in other couples	0.91	0.25	–0.61	–1.99	–0.10
– Women in other couples	0.70	–0.22	–0.63	–6.55	–2.94
– Single parents youngest child 0–17	0.68	0.24	–0.81	0.23	–0.59
– Singles	0.86	0.09	–0.28	–0.56	–0.18
Labour productivity per hour		0.18	0.27	0.61	0.36
Gini coefficient[f]	0.24	5.18	0.00	–7.95	0.76

[a] Knock-on effects are not presented for budgetary neutral scenarios, as these effects are expressed as a percentage of the initial budgetary impulse which equals zero in the first four scenarios. [b] A flat tax rate of 39.7%. [c] A flat tax rate of 45.3% and a lump-sum benefit of 1.950 euros for all adults. [d] An introduction of an unconditional basic income of 50% of the net welfare benefit (8,248 euros) and a flat tax rate of 56.5%. [e] Joint taxation, financed through an increase of 1.9 percentage points in the tax rates in the four tax brackets. [f] Gini coefficient of disposable household income, using equivalence scales. The Gini coefficient is calculated over the full Dutch adult population with a gross income of over 66% of the annual minimum wage.

increases marginal and participation tax rates at the lower end of the income distribution, which is more responsive to tax changes in terms of hours worked than the upper end of the income distribution.[42]

Column (3) shows that generic income support via the introduction of an unconditional basic income has a strong adverse effect on labour force participation. We simulate a basic income of 50% of the net welfare benefit level. All adults qualify for this basic income. For adults receiving social benefits (e.g. welfare, unemployment, disability or retirement benefits) we reduce the benefit level, so that together with the basic income their disposable income does not change. We finance the basic income scenario by abolishing the general in-work tax credit for all workers, which is in line with the idea that this type of income support is unconditional, and a flat tax rate of 56.6%.[43] The flat tax rate of 56.6% implies a considerable increase in all marginal tax rates, in particular for adults in the lower tax brackets. Furthermore, the introduction of a basic income increases income for non-working partners. Labour supply decreases both on the extensive and intensive margin, and in total by –5.5%.[44] On the upside, inequality decreases as well, by almost 8%.

Finally, column (4) simulates a move from individual to joint taxation, as in the tax systems of e.g. France, Germany and the US.[45] We simulate joint taxation by taking the sum of taxable income from both partners and then assign half of the total taxable household income to both partners. We finance this scenario by increasing marginal tax rates in all four tax brackets by 1.9 percentage points. Total labour supply decreases by –2.2%. Most

[42]Simulation studies for other countries also find that a flat tax may stimulate labour supply, but only when marginal tax rates and basic allowances are set relatively low (Aaberge et al., 2000; Fuest et al., 2008).

[43]When the general in-work tax credit is not abolished, we still obtain a strong negative effect on labour supply.

[44]Clavet et al. (2013) and Horstschraer et al. (2010) also use a structural labour supply model to study the labour supply effects of a basic income, for Quebec (Canada) and Germany, respectively. Both studies find a strong decrease in total hours worked. Our results are also consistent with the findings from the basic income experiments in the US and Canada. (Hum and Simpson, 1993, Table 2) gives an overview. For the United States, hours worked by men in couples, by women in couples, and by single women declined by –5% to –7%, –17% to –21% and –13% to –17%, respectively. For Canada, hours worked by men in couples, by women in couples, and by single women declined by a respective –1%, –3% and –7%.

[45]In the Netherlands, the Reformed Political Party (SGP), in their election manifesto, proposed a tax system which would allow incomes to be split.

women in couples are secondary earners and face a relatively low marginal tax rate under the tax system in the baseline. Joint taxation means that the marginal tax rate increases for secondary earners and they reduce their labour supply. The effective marginal tax rate for primary earners decreases, and they increase their labour supply, but to a much lesser extent. Furthermore, inequality increases. Hence, this scenario scores unfavourably both in terms of hours worked and in income inequality.[46]

8 Discussion and conclusion

In this paper we used both structural models and quasi-experimental studies to study the effectiveness of tax-benefit reforms. Using a very large and rich data set, we estimate structural discrete-choice models for a large number of household types. We uncover large differences in the labour supply responses between various demographic groups, mostly related to the age of the youngest child. We also find that the decision of whether or not to participate is more responsive to financial incentives than the hours-per-week decision, although the hours-per-week decision is still non-negligible for women in couples with children. We used the structural model to simulate a number of key reforms from the past, and compared the predictions of the structural model with the outcomes of quasi-experimental studies on the same reforms. We find that the structural model predicts the estimated treatment effects from the quasi-experimental studies rather well.

We then conduct a counterfactual policy analysis with the structural model, and study the effectiveness of potential tax-benefit reforms in stimulating labour supply. We find that reducing marginal tax rates is not an effective way to promote labour supply. In-work benefits targeted at low-wage earners are more effective. Policies targeted at working mothers with young children generate the largest labour supply response, but generate little additional

[46]Kabatek et al. (2014) estimate a discrete-choice labour supply model for French couples to evaluate a shift from joint taxation to individual taxation. They find that a system with joint taxation discourages female labour supply. Callan et al. (2009) estimate a discrete-choice model for couples, using Irish data. They find that a shift from joint taxation to individual taxation stimulates women's labour supply.

revenue for the government. With the structural model we also simulate some major tax reforms that feature prominently in the current policy debate. We find that proposals for a move to a flat tax system, a basic income system or a system with joint taxation are not effective in stimulating labour supply. Indeed, an efficient tax system accounts for the large heterogeneity in responses, between different demographic groups (e.g. primary vs. secondary earners and with vs. without young children) and different decision margins (e.g. extensive vs. intensive margin), and therefore cannot be too simple.

Although we believe that our analysis makes a number of improvements over previous studies on the effectiveness of tax-benefit reforms, it still has a number of limitations. We ignore involuntary unemployment (and a potential difference between preferred and actual working hours). However, estimating a double-hurdle model (Cragg, 1971), we find that accounting for involuntary unemployment makes little difference in the employment responses to changes in financial incentives (De Boer, 2018).[47] Furthermore, we ignore responses to marginal (and participation) tax rates other than labour supply. Part of the modern literature on public finance looks at a broader range of behavioural responses, by considering the so-called elasticity of taxable income, see Saez et al. (2012) for an overview. Indeed, a recent study by Jongen and Stoel (2019) for the Netherlands shows that the elasticity-of-taxable-income may be higher than the labour supply elasticity, suggesting larger distortions from tax rates than by looking solely at labour supply. We further ignore the life cycle. A number of studies have shown that accounting for life-cycle effects can be important for the analysis of tax-benefit reform (e.g. Blundell et al., 2016; Imai and Keane, 2004; Keane, 2011). This would be an interesting direction for future research. However, the data set we used does not include data on, for example, consumption or savings, which makes it difficult to estimate a life-cycle model, and it should be noted that there is often a trade-off in modelling different

[47]Regarding the difference between actual and preferred hours of work, we do not have data on preferred hours of work in our data set. However, this seems to be a much smaller problem in the Netherlands than in many other OECD countries. For example, OECD (2013) reports that just 5% of women working part-time in the Netherlands would like to work more hours, compared to, for example, 13% in Germany, 28% in France and 55% in Spain.

parts of economic behaviour, due to the numerical complexities that arise.[48] Finally, we assume that all people are fully aware of their budget constraint. However, recent work by Chetty et al. (2009) shows that information, or the lack thereof, can play an important role in the behavioural responses to financial incentives. This too seems an interesting direction for future research.[49]

Compliance with Ethical Standards:

The authors declare that they have no conflict of interest.

References

Aaberge, R., Colombino, U., and Strom, S. (1999). Labour supply in Italy: An empirical analysis of joint household decisions, with taxes and quantity constraints. *Journal of Applied Econometrics*, 14(4):403–422.

Aaberge, R., Colombino, U., and Strom, S. (2000). Labor supply responses and welfare effects from replacing current tax rules by a flat tax: Empirical evidence from Italy, Norway and Sweden. *Journal of Population Economics*, 13(4):595–621.

Aaberge, R., Dagsvik, J., and Strom, S. (1995). Labor supply responses and welfare effects of tax reforms. *Scandinavian Journal of Economics*, 97(4):635–659.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

Attanasio, O., Meghir, C., and Santiago, A. (2011). Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA. *Review of Economic Studies*, 79(1):37–66.

Bargain, O. and Doorley, K. (2017). The effect of social benefits on youth unemployment: Combining regression discontinuity and a behavioral model. *Journal of Human Resources*, 52(4):1032–1059.

Bargain, O., Orsini, K., and Peichl, A. (2014). Comparing labor supply elasticities in Europe and the United States: New results. *Journal of Human Resources*, 49(3):723–838.

Bargain, O. and Peichl, A. (2017). Steady-state labor supply elasticities: A survey. *IZA Journal of Labor Economics*, forthcoming.

Bettendorf, L., Folmer, K., and Jongen, E. (2014). The dog that did not bark: The EITC for single mothers in the Netherlands. *Journal of Public Economics*, 119:49–60.

[48] Estimating the preferences for all subgroups in our static model already took several weeks.

[49] See e.g. Bosch et al. (2019) for a bunching analysis of high effective marginal tax rates resulting from income support in the Netherlands, which suggests rather limited bunching, which may be due to inattention or frictions.

Bettendorf, L., Jongen, E., and Muller, P. (2015). Childcare subsidies and labour supply - Evidence from a large Dutch reform. *Labour Economics*, 36:112–123.

Blau, D. (2003). Child care subsidy programs. In Moffitt, R., editor, *Means-Tested Transfer Programs in the United States*, pages 443–516. NBER, Cambridge.

Blau, D. and Kahn, L. (2007). Changes in the labor supply behavior of married women: 1980-2000. *Journal of Labor Economics*, 25(3):393–438.

Blundell, R., Chiappori, P.-A., and Meghir, C. (2007). Collective labour supply: Heterogeneity and non-participation. *Review of Economic Studies*, 74:417–455.

Blundell, R., Costa-Dias, M., Meghir, C., and Shaw, J. (2016). Female labour supply, human capital and welfare reform. *Econometrica*, 84(5):1705–1763.

Blundell, R., Duncan, A., McCrae, J., and Meghir, C. (2000). The labour market impact of the Working Families' Tax Credit. *Fiscal Studies*, 21(1):75–104.

Blundell, R., Duncan, A., and Meghir, C. (1998). Estimating labor supply responses using tax reforms. *Econometrica*, 66:827–861.

Blundell, R. and MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, chapter 27, pages 1559–1695. Elsevier, Amsterdam.

Blundell, R. and Shephard, A. (2012). Employment, hours of work and the optimal taxation of low income families. *Review of Economic Studies*, 79(2):481–510.

Bosch, N. and Jongen, E. (2013). Intensive margin responses when workers are free the choose: Evidence from a Dutch tax reform. Paper presented at EALE 2013.

Bosch, N., Jongen, E., Leenders, W., and Mohlmann, J. (2019). Non-bunching at kinks and notches in cash tranfers in the Netherlands. *International Tax and Public Finance*, forthcoming.

Bosch, N. and van der Klaauw, B. (2012). Analyzing female labor supply - Evidence from a Dutch tax reform. *Labour Economics*, 19:271–280.

Brewer, M., Duncan, A., Shephard, A., and Suarez, M. (2006a). Did the Working Families' Tax Credit work? The impact of in-work support on labour supply in Great Britain. *Labour Economics*, 13:699–720.

Brewer, M., Duncan, A., Shephard, A., and Suarez, M. (2006b). Did the Working Families' Tax Credit work? The impact of in-work support on labour supply in Great Britain. *Labour Economics*, 13:699–720.

Brewer, M., Saez, E., and Shephard, A. (2010). Means-testing and tax rates on earnings. In Mirrlees, J. A., Adam, S., Besley, T. J., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. D., and Poterba, J. M., editors, *The Mirrlees Review – Dimensions of Tax Design*, chapter 3, pages 202–274. Oxford University Press, Oxford.

Cai, L., Kalb, G., Tensg, Y.-P., and Vu, H. (2008). The effect of financial incentives on labour supply: Evidence for lone parents from microsimulation and quasi-experimental evaluation. *Fiscal Studies*, 29(2):285–325.

Callan, T., Van Soest, A., and Walsh, J. (2009). Tax structure and female labour supply: Evidence from Ireland. *Labour*, 23(1):1–35.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1:451–488.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177.

Clavet, N., Duclos, J., and Lacroix, G. (2013). Fighting poverty: Assessing the effect of guaranteed minimum income proposals in Quebec. *Canadian Public Policy*, 39(4):491–516.

CPB (2015). *Kansrijk Arbeidsmarktbeleid*.

Cragg, J. (1971). Some statistical models for limited dependent variables with applications to the demand for durable goods. *Econometrica*, 39:829–844.

De Boer, H.-W. (2018). A structural analysis of labour supply and involuntary unemployment in the Netherlands. *De Economist*, 166(3):285–308.

De Boer, H.-W., Jongen, E., and Kabatek, J. (2015). The effectiveness of fiscal stimuli for working parents. IZA Discussion Paper 9298, Bonn.

Deaton, A. and Paxson, C. (1994). Intertemporal choice and inequality. *Journal of Political Economy*, 102(3):437–467.

Ericson, P. and Flood, L. (2012). A microsimulation approach to an optimal Swedish income tax. *International Journal of Microsimulation*, 5(2):2–21.

European Union (2014). *Taxation Trends in the European Union, 2014 edition*. European Union, Brussels.

Euwals, R., Knoef, M., and van Vuuren, D. (2011). The trend in female labour force participation: What can be expected for the future? *Empirical Economics*, 40:729–753.

Fuest, C., Peichl, A., and Schaefer, T. (2008). Is a flat tax reform feasible in a grown-up democracy of Western Europe? A simulation study for Germany. *International Tax and Public Finance*, 15:620–636.

Geyer, J., Haan, P., and Wrohlich, K. (2015). The effects of family policy on mothers' labor supply: Combining evidence from a structural model and a quasi-experimental approach. *Labour Economics*, 36(C):84–98.

Hansen, J. and Liu, X. (2015). Estimating labor supply responses and welfare participation: Using a natural experiment to validate a structural model. *Canadian Journal of Economics*, 48(5):1831–1854.

Heckman, J. (1979). Sample selection bias as a specifcation error. *Econometrica*, 47(1):153–161.

Heckman, J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, 48(2):356–398.

Heim, B. (2007). The incredible shrinking elasticities: Married female labor supply, 1978-2002. *Journal of Human Resources*, 42(4):881–918.

Horstschraer, J., M., C., and Schnabel, R. (2010). An unconditional basic income in the family context: Labor supply and distributional effects. ZEW Discussion Paper No. 10-091, Munich.

Hum, D. and Simpson, W. (1993). Economic response to a guaranteed annual income: Experience from Canada and the United States. *Journal of Labor Economics*, 11(1):S263–S296.

Imai, S. and Keane, M. (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review*, 45(2):601–641.

Jacobs, B., De Mooij, R., and Folmer, K. (2009). Analyzing a flat tax in the Netherlands. *Applied Economics*, 42(25):3209–3320.

Jongen, E. and Stoel, M. (2019). Estimating the elasticity of taxable labour income in the Netherlands. *De Economist*, forthcoming.

Kabatek, J., Van Soest, A., and Stancanelli, E. (2014). Income taxation, labour supply, and housework: A discrete choice model for French couples. *Labour Economics*, 27:30–43.

Keane, M. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, 156:3–20.

Keane, M. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, 49(4):961–1075.

Keane, M. and Moffitt, R. (1998). A structural model of multiple welfare program participation and labor supply. *International Economic Review*, 39(3):553–589.

Koot, P., Vlekke, M., Berkhout, E., and Euwals, R. (2016). MIMOSI: Microsimulatiemodel voor belastingen, sociale zekerheid, loonkosten en koopkracht. CPB Background Document, The Hague.

Lokshin, M. (2004). Household childcare choices and women's work behavior in Russia. *Journal of Human Resources*, 39(4):1094–1115.

Lundberg, S. (1985). The added worker effect. *Journal of Labor Economics*, 3(1):11–37.

MaCurdy, T., Green, P., and Paarsch, H. (1990). Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources*, 25:415–490.

McFadden, D. (1978). Modeling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North-Holland, Amsterdam.

Ministry of Finance (2015). Belastingherziening. Ministry of Finance, The Hague.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

OECD (2013). *OECD Labour Force Statistics 2013*. OECD, Paris.

OECD (2016). *OECD Labour Force Statistics 2016*. OECD, Paris.

Pronzato, C. (2018). Comparing quasi-experimental designs and structural models for policy evaluation: The case of a reform of lone parental welfare. *CESifo Economic Studies*, page forthcoming.

Quist, A. (2015). Marginale druk en participatiebelasting per huishoudtype in 2015. CPB Background Document, The Hague.

Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics*, 117(3):1039–1073.

Saez, E., Slemrod, J., and Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 2012*, 50:1:350.

Statistics Netherlands (2012). Documentatierapport Arbeidsmarktpanel 1999-2009V1. Statistics Netherlands, Leidschenveen.

Thoresen, T. and Vatto, T. (2015). Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature. *Labour Economics*, 37(C):38–53.

Todd, P. and Wolpin, K. (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, 96(5):1384–1417.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

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

Van Soest, A. and Das, M. (2001). Family labour supply and proposed tax reforms in the Netherlands. *De Economist*, 149:191–218.

Van Soest, A., Das, M., and Gong, X. (2002). A structural labor supply model with flexible preferences. *Journal of Econometrics*, 107(1-2):345–374.

Vella, F. and Verbeek, M. (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics*, 13:163–183.

Vella, F. and Verbeek, M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics*, 90(2):239–263.

Supplementary Material to
"Analysing Tax-Benefit Reforms in the Netherlands Using Structural Models and Natural Experiments"

Henk-Wim de Boer and Egbert L.W. Jongen

September 2019

A Effective marginal and participation tax rates

The financial incentives implicit in the tax-benefit system are illustrated using the so-called effective marginal tax rates (EMTRs) and participation tax rates (PTRs). We use the tax-benefit system of 2015 as the baseline. Panels (a) and (b) show the incentives for singles, without and with children respectively. Panels (c) and (d) show the financial incentives for single earners, without and with children respectively. Single earners are working individuals in couples with a non-working partner. Panels (e) and (f) give the incentives for primary earners, without and with children respectively. Primary earners are working individuals in two-earner couples who have a higher income than their working partner. Finally, panels (g) and (h) show the incentives for secondary earners, again without and with children respectively. Secondary earners are working individuals in two-earner couples who have a lower income than their working partner.[50]

First consider the EMTRs. To calculate the EMTRs, we first calculate disposable household income (y_0) at the initial gross labour income level (z_0). Next, we increase gross labour income by 3% (z_1), and calculate the corresponding new disposable household income (y_1). The EMTRs are then calculated as 1 minus the change in disposable household income over the change is gross labour income:

$$EMTR = 1 - \left(\frac{y_1 - y_0}{z_1 - z_0} \right). \tag{A.1}$$

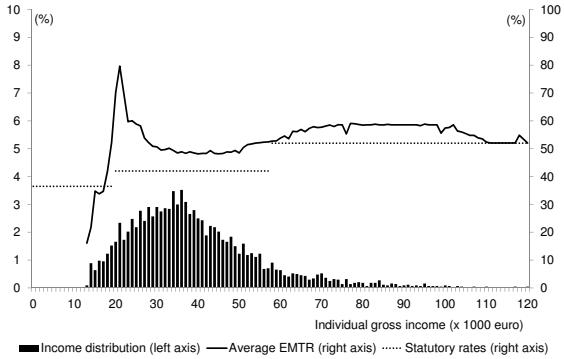
Following this procedure, the EMTRs account for the statutory tax rates and the income dependent tax credits and subsidies. The dotted lines in Figure A.1 show the statutory rates. In 2015 the Netherlands had four tax brackets for personal labour income, with the statutory rate being the same for the second and third bracket.[51]

[50] For each group, the average per income bin of 1000 euros is shown. A decomposition of the EMTRs and PTRs in the different income-dependent elements can be found in Quist (2015).

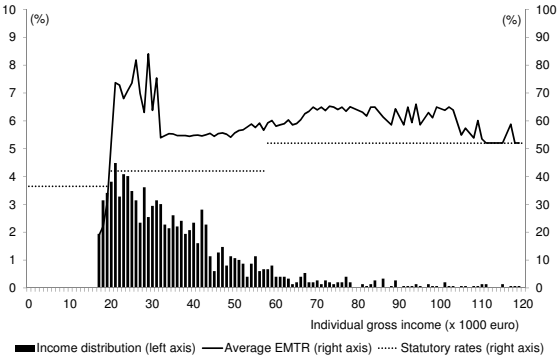
[51] In 2015, the tax rate in the first, second, third and fourth (open) tax bracket was 36.5, 42, 42 and 52%,

Figure A.1: Effective marginal tax rates per household type in the Netherlands (2015)

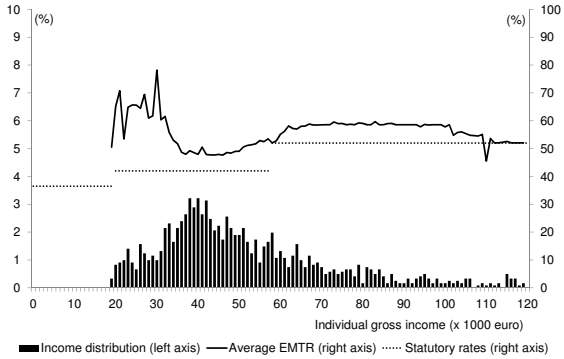
(a) Singles without children



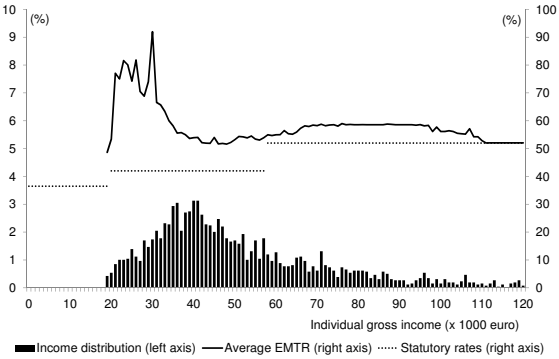
(b) Singles with children



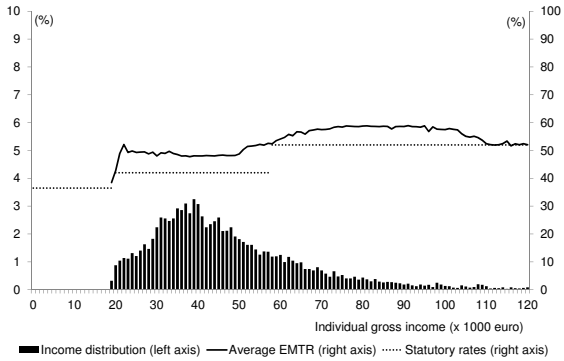
(c) Single earners without children



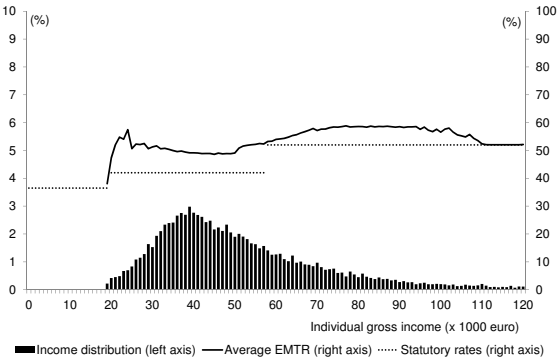
(d) Single earners with children



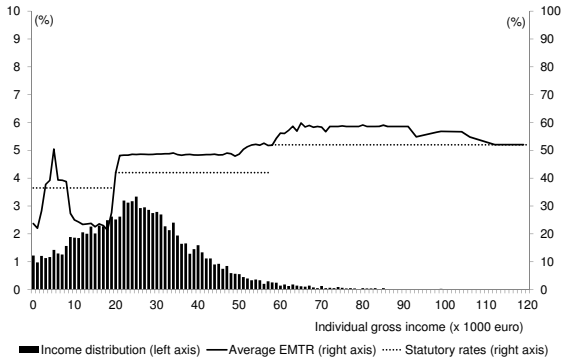
(e) Primary earners without children



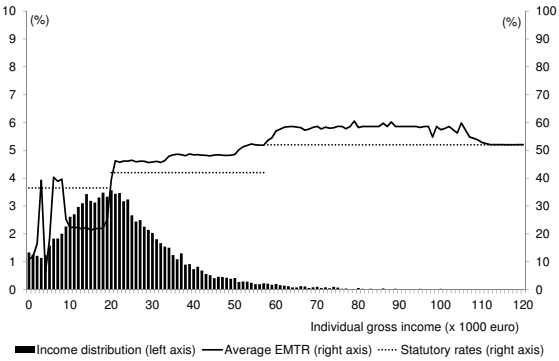
(f) Primary earners with children



(g) Secondary earners without children



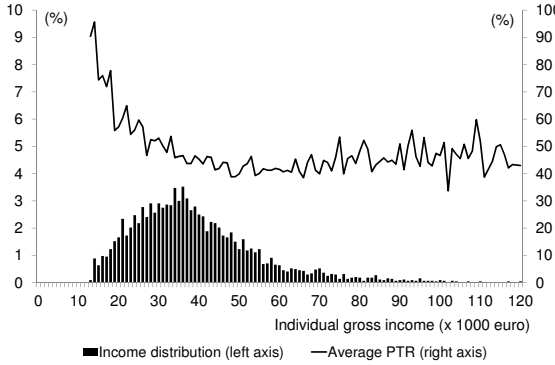
(h) Secondary earners with children



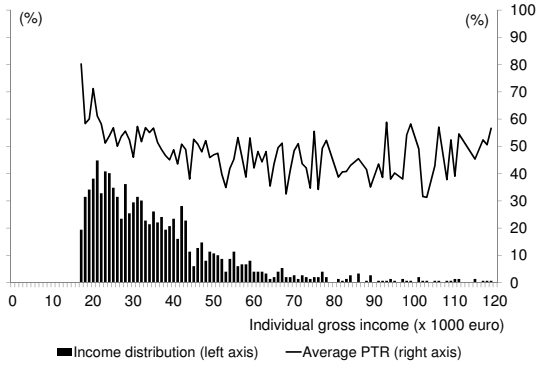
Source: Quist (2015).

Figure A.2: Participation tax rates per household type in the Netherlands (2015)

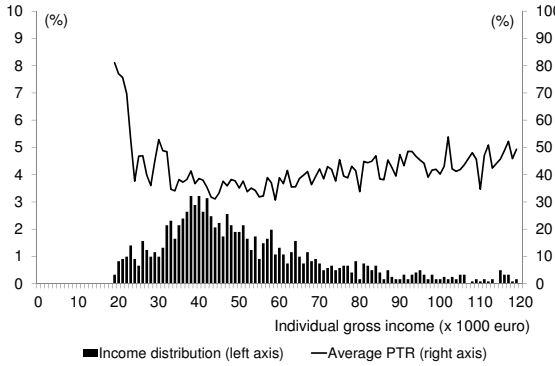
(a) Singles without children



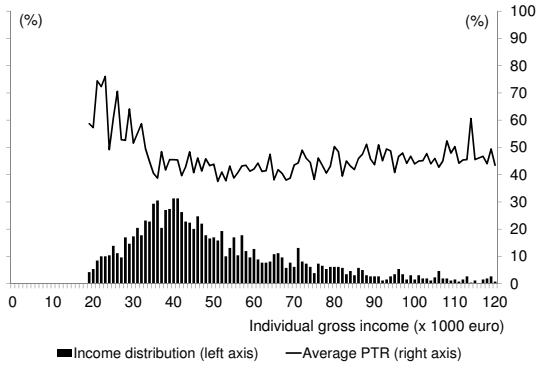
(b) Singles with children



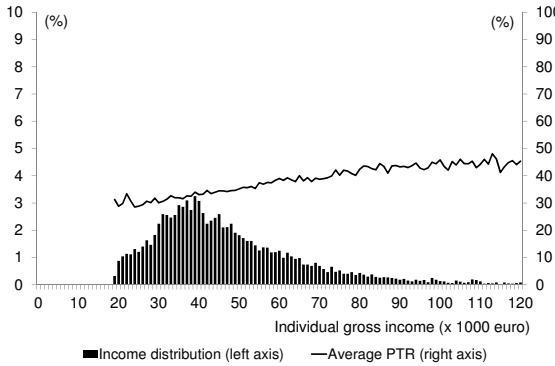
(c) Single earners without children



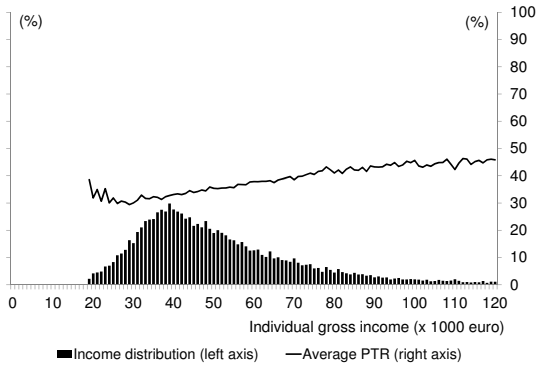
(d) Single earners with children



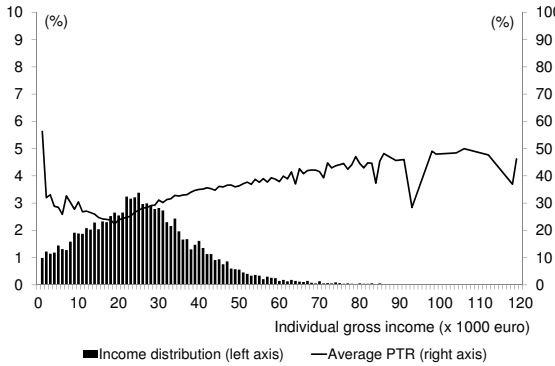
(e) Primary earners without children



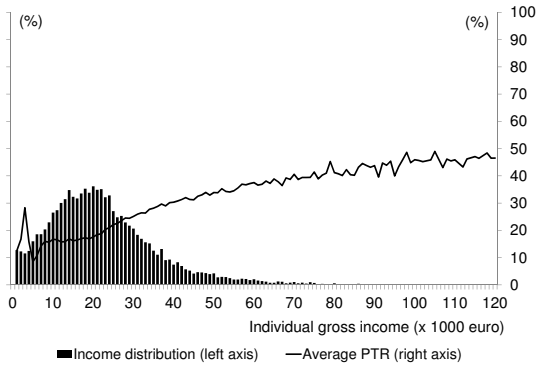
(f) Primary earners with children



(g) Secondary earners without children



(h) Secondary earners with children



Source: Quist (2015).

For singles with a gross labour income below the full-time annual minimum wage of approximately 18,000 euro, EMTRs are somewhat lower than the statutory rates, due to the phase-in of the general earned income tax credit (between 0 and 20 thousand euro). However, EMTRs are substantially higher for singles with an annual income between the minimum wage and the modal wage of approximately 35,000 euros because of the phase-out of rent and health care subsidies (see Quist, 2015). Figure A.1 also gives the density of workers at each income level, which shows a large number of workers between the minimum wage and the modal wage facing these high EMTRs. EMTRs remain higher than statutory rates also beyond the modal wage, because of the phase-out of the general tax credit (for annual wages between 20,000 euros and 57,000 euros) and the general earned income tax credit (between 50,000 and 101,000 euros). EMTRs are still higher for singles with children than for singles without children. This is due to the phase-out of targeted income support for households with dependent children such as the income-dependent child benefit (see Quist, 2015). EMTRs for single earners are also substantially higher than the statutory rates, mainly because of the phase-out of rent subsidies, health care subsidies, the general tax credit and the earned income tax credit. The phase-out of subsidies for families with children increases EMTRs further for single earners with children. However, because many of the single earners earn a relatively high income, a smaller number of them face the relatively high EMTRs between the minimum wage and the modal wage. EMTRs for primary earners are generally not that much higher than statutory rates. Because they have a working partner, they typically do not qualify for rent subsidies and/or health care subsidies. Also, primary earners with children are less likely to qualify for the targeted income support for families with children, because they have a working partner. Finally, secondary earners often have an EMTR that is lower than the statutory rate. A large number of them has an income below the minimum wage, where the (general) earned income tax credit is phased in. Furthermore, secondary earners with children receive an additional earned income tax credit, which also has a phase-in range.[52] This further lowers their EMTRs. Working in the opposite direction is the phase-out of the general tax credit (and the general earned income tax credit, but secondary earners typically do not earn enough to be affected by this phase-out).[53]

Figure A.2 gives the PTRs. To calculate the PTRs we first calculate disposable household

respectively.

[52]This tax credit jumps to 1,033 euros at a personal gross wage income of 4,857 euros and is subsequently phased in at a rate of 4%, until the maximum amount of 2,152 euros is reached at a personal gross wage income of almost 33,000 euros.

[53]Here we follow Quist (2015) and do not include childcare subsidies in the EMTRs, because their effect on the EMTRs depends on what one assumes about the additional use of childcare to generate the 3% additional income. However, in the empirical analysis in the main text we explicitly account for the use of childcare, as well as the associated costs and subsidies.

income when the individual does not work (y_{work}). Next, we calculate disposable household income when the individual does work (y_{no_work}). The PTRs are then calculated as 1 minus the difference in disposable household income over the difference in gross labour income (z_{work})[54]:

$$PTR = 1 - \left(\frac{y_{work} - y_{no\_work}}{z_{work}} \right). \tag{A.2}$$

Following this procedure, the PTRs account for taxes paid and subsidies received when the individual is working, and the withdrawal of welfare benefits when they do not (provided they would qualify for welfare benefits, which for example depends on income of a partner).[55]

The resulting PTRs are relatively high for singles without children with a relatively low income, primarily because of the withdrawal of benefits.[56] At higher income levels, PTRs level out, as the withdrawal of benefits becomes less important as a percentage of gross wage income. The profile of PTRs for single parents is actually not that different from the PTRs for singles. Non-working single parents qualify for additional income support, but so do working single parents. However, note that the region with high PTRs is more densely populated for single parents than for singles without children, because single parents are more likely to work part-time. PTRs for single earners are also higher for low incomes, but the density in the income interval with high PTRs is much lower for this group. PTRs are relatively low for primary and secondary earners, because they typically do not qualify for benefits when they do not work (because of the income of the partner). Finally, PTRs are particularly low for secondary earners with children, because they qualify for additional in-work benefits.

[54] Gross labour income is zero when the individual does not work.

[55] Quist (2015) only calculates the PTRs for workers. When non-workers have higher PTRs for a given level of income than workers, PTRs will be higher for workers and non-workers combined.

[56] See Quist (2015) for a decomposition of PTRs by individual items.

B Wage equations

For the employed we use observed wages. For the non-employed we simulate wages, accounting for potential selection bias (Heckman, 1979). To this end, we run wage regressions by gender and then by level of education, where education is split into three levels (lower, middle and higher educated).

We use panel data techniques to account for unobserved individual-specific effects. We perform a Hausman test in order to test whether random effects or fixed effects are appropriate. For all groups, we reject the null hypothesis that the individual-specific effects are uncorrelated with the regressor and therefore we prefer fixed effects over random effects estimation. However, we lose information on time-invariant regressors with fixed effects and therefore opt for the quasi-fixed effects model (Mundlak, 1978).

To account for the possibility of selection we first estimate the probability of participation using a pooled probit regression:

$$p_{it} = x'_{it}\gamma + z'_{it}\theta + \nu_{it}, \tag{B.1}$$

where vector z_{it} contains variables that are expected to have an effect on the probability of participation but not on wages (an exclusion restriction). From this regression we determine the inverse Mills' ratio:

$$invmills_{it} = \phi(p_{it})/\Phi(p_{it}). \tag{B.2}$$

The inverse Mills' ratio is then included in the quasi-fixed effects model:

$$ln(w_{it}) = x'_{it}\beta + \omega_i + \bar{x_i}'\pi + \lambda_t invmills_{it} + \epsilon_{it}, \tag{B.3}$$

where the individual specific effect consists of a random part, ω_i with $\sim IID(0, \sigma^2_\omega)$, and a part which is allowed to be correlated with regressors $\bar{x_i}'\pi$. Here, $\bar{x_i}$ is the average of time-varying variables, such as age. A significant coefficient for an element of π provides evidence that the individual specific effect is correlated with one of the regressors.

Tables B.1 and B.2 show estimation results for men and women in couples and single men and women, respectively. We estimate the wage equations separately for men and women, and by level of education. We use age splines since we expect that the relationship between wage and age is nonlinear. The results show that wage increases with age but at a diminishing rate. This is in line with other studies (Vella and Verbeek, 1998, 1999). For both singles and couples we see that the age profile is steeper for higher educated individuals. We also include cohort and year dummies in the regression. Because of perfect collinearity between age, cohort and period we use transformed time dummies, following Deaton and Paxson

Table B.1: Wage equations couples

	Men			Women		
	Lower educ.	Middle educ.	Higher educ.	Lower educ.	Middle educ.	Higher educ.
Age						
18-30	0.042***	0.052***	0.078***	0.037***	0.042***	0.052***
31-40	0.021***	0.032***	0.048***	0.023***	0.025***	0.035***
41-50	0.014***	0.021***	0.029***	0.022***	0.020***	0.024***
51-63	0.007***	0.005***	0.011***	0.024***	0.019***	0.008***
Cohort[a]						
1980-1989	0.127***	0.189***	0.210***	0.158***	0.167***	0.217***
1975-1980	0.038**	0.112***	0.149***	0.069***	0.089***	0.153***
1970-1975	0.020**	0.055***	0.095***	0.026***	0.053***	0.099***
1960-1965	-0.003	-0.019***	-0.019***	-0.004	-0.014**	-0.054***
1955-1960	-0.017	-0.035***	-0.058***	0.006	-0.021**	-0.058***
<1955	0.011	0.01	-0.018	0.009	-0.013	-0.019
Year						
2006	0.004	0.005	0.003	0.006	0.004	0.002
2007	-0.005	-0.007	-0.004	-0.008	-0.004	-0.002
2008	-0.002***	-0.002***	-0.003***	-0.003***	-0.003***	-0.003***
2009	0.003***	0.003***	0.003***	0.005***	0.004***	0.002***
Ethnicitity[a]						
Western immigrant	-0.043***	-0.048***	-0.013*	-0.003	-0.036***	-0.067***
Non-western immigrant	-0.207***	-0.236***	-0.294***	-0.079***	-0.111***	-0.194***
Partner						
married	0.025***	0.013***	0.025***	0.001	-0.009***	-0.020***
Mundlak age averages						
18-30	0.001	-0.004	-0.004	-0.005	-0.003	0.002
31-40	-0.006***	-0.003**	0.000	-0.012***	-0.008***	-0.004***
41-50	-0.009***	-0.010***	-0.014	-0.018***	-0.013***	-0.014***
51-63	-0.011***	-0.018***	-0.021	-0.022***	-0.025***	-0.023***
Province[a]						
Groningen	-0.126***	-0.138***	-0.116***	-0.041***	-0.055***	-0.055***
Friesland	-0.114***	-0.138***	-0.119***	-0.060***	-0.075***	-0.077***
Drenthe	-0.110***	-0.115***	-0.097***	-0.058***	-0.050***	-0.063***
Overijssel	-0.094***	-0.109***	-0.095***	-0.058***	-0.063***	-0.065***
Flevoland	-0.014	-0.033***	-0.064***	-0.046***	-0.037***	-0.058***
Gelderland	-0.063***	-0.065***	-0.066***	-0.049***	-0.043***	-0.050***
Noord-Brabant	-0.055***	-0.050***	-0.081***	-0.031***	-0.058***	-0.070***
Zeeland	-0.040***	-0.067***	-0.051***	-0.056***	-0.051***	-0.055***
Limburg	-0.116***	-0.130***	-0.107***	-0.073***	-0.075***	-0.064***
Urban area						
>150.000 habitants	-0.022***	-0.006**	-0.004	0.013**	0.016**	0.031**
Inverse Mills	0.000	0.120**	0.082	-0.041**	0.010	0.192**
Constant	1.320***	1.140**	0.555**	1.395**	1.311**	1.051**
Observations	106,649	209,769	167,390	77,924	195,231	130,975

*** p<0.01, ** p<0.05, * p<0.1. [a] Reference group: born in 1965–1970, autochtonous and living in the Western region of the Netherlands.

Table B.2: Wage equations singles

	Men			Women		
	Lower educ.	Middle educ.	Higher educ.	Lower educ.	Middle educ.	Higher educ.
Age						
18-30	0.068***	0.063***	0.077***	0.066***	0.058***	0.060***
31-40	0.018***	0.031***	0.050***	0.021***	0.028***	0.041***
41-50	0.010***	0.018***	0.026***	0.025***	0.020***	0.023***
51-63	0.011***	0.015***	0.016***	0.025***	0.023***	0.018***
Cohort[a]						
1980-1989	0.207***	0.229***	0.158***	0.265***	0.221***	0.215***
1975-1980	0.026	0.107***	0.073***	0.046	0.100***	0.149***
1970-1975	0.014	0.032***	0.049***	0.015	0.034***	0.089***
1960-1965	0.013	0.015	-0.025	-0.005	-0.024**	-0.041***
1955-1960	0.035	0.045**	-0.058**	-0.022	-0.024	-0.070***
<1955	-0.003	0.002	-0.031	-0.012	-0.028**	-0.037**
Year						
2006	0.005	0.005	0.005	0.009	0.006	0.001
2007	-0.007	-0.006	-0.007	-0.011	-0.006	-0.002
2008	0.000	-0.003***	-0.002**	-0.003**	-0.005***	-0.001
2009	0.003**	0.004***	0.004***	0.006***	0.005***	0.001*
Ethnicitity[a]						
Western immigrant	-0.011	-0.004	0.009	-0.017*	-0.025***	-0.025**
Non-western immigrant	-0.063***	-0.078***	-0.166***	-0.048***	-0.057***	-0.111***
Mundlak age averages						
18-30	-0.011***	-0.002	-0.008	-0.014***	-0.003	-0.002
31-40	-0.009**	-0.009***	-0.016***	-0.015***	-0.014***	-0.007**
41-50	-0.009***	-0.014***	-0.013***	-0.017***	-0.012***	-0.013***
51-63	-0.006*	-0.008***	-0.011***	-0.018***	-0.020***	-0.019***
Province[a]						
Groningen	-0.062***	-0.083***	-0.086***	-0.044***	-0.042***	-0.070***
Friesland	-0.070***	-0.120***	-0.119***	-0.066***	-0.084***	-0.086***
Drenthe	-0.059***	-0.114***	-0.070***	-0.075***	-0.078***	-0.078***
Overijssel	-0.074***	-0.094***	-0.073***	-0.064***	-0.064***	-0.055***
Flevoland	-0.033*	-0.015	-0.006	-0.030**	-0.050***	-0.038**
Gelderland	-0.059***	-0.068***	-0.062***	-0.054***	-0.047***	-0.050***
Noord-Brabant	-0.021	-0.051***	0.020	-0.049***	-0.050***	-0.068***
Zeeland	-0.038***	-0.071***	-0.043***	-0.064***	-0.057***	-0.056***
Limburg	-0.090***	-0.097***	-0.079***	-0.068***	-0.069***	-0.059***
Urban area						
>150.000 habitants	-0.012	0.019***	0.021***	0.033***	0.034***	0.019***
Inverse Mills	-0.112	-0.296***	-0.325***	-0.100***	-0.201***	-0.213***
Constant	0.798***	0.750***	0.720***	0.829***	0.877***	0.951***
Observations	33,152	50,679	32,463	28,804	54,476	38,254

*** p<0.01, ** p<0.05, * p<0.1. [a] Reference group: born in 1965–1970, autochtonous and living in the Western region of the Netherlands.

Table B.3: Exclusion restrictions couples

	Men			Women		
	Lower educ.	Middle educ.	Higher educ.	Lower educ.	Middle educ.	Higher educ.
Youngest child 0–3	-0.195***	-0.125***	-0.103***	-0.626***	-0.426***	-0.331***
Youngest child 4–11	-0.139***	-0.082***	-0.068***	-0.397***	-0.360***	-0.245***
Youngest child 12-17	-0.098***	-0.011	0.027	-0.171***	-0.171***	-0.050**
Lower educ. partner	0.017	0.062***	0.002	-0.117***	0.014	0.122***
Middle educ. partner	0.091***	0.060***	0.045***	-0.001	0.096***	0.173***
Wald test	81.860***	69.680***	41.860***	746.210***	758.870***	374.660***

*** p<0.01, ** p<0.05, * p<0.1.

Table B.4: Exclusion restrictions singles

	Men			Women		
	Lower educ.	Middle educ.	Higher educ.	Lower educ.	Middle educ.	Higher educ.
Youngest child 0–3	-0.608***	-0.128	-0.522**	-1.196***	-0.924***	-0.566***
Youngest child 4–11	-0.264***	-0.078	0.311**	-0.665***	-0.545***	-0.297***
Youngest child 12-17	-0.033	0.219***	0.491***	-0.206***	-0.182***	-0.076
Wald test	20.790***	11.840***	40.520***	979.400***	539.480***	68.190***

*** p<0.01, ** p<0.05, * p<0.1.

(1994). The time dummies for 2006 and 2007 depend on the dummies for later years and are calculated manually.[57] Year dummies are significant in most specifications while the cohort variables are jointly significant for most subgroups. Wages are lower on average for non-Western immigrants. The coefficients for the Mundlak age averages are jointly significant in all specifications, but have no economic interpretation.

The inverse Mills' ratio is significant for most groups, indicating that selection bias is present for most groups.[58] Tables B.3 and B.4 show the exclusion restrictions for couples and singles, respectively. As exclusion restrictions, we use the presence of young children in the participation equation, following e.g. Van Soest (1995) and Bargain et al. (2014). For couples, we also use the education level of the partner, following Blundell et al. (2007). The exclusion restrictions turn out te be jointly significant, as indicated by the Wald test.

[57] t2006=-(d2007+d2008+d2009) and t2007=-2*d2008-3*d2009

[58] We also include an attrition indicator in order to test for the presence of attrition bias (not shown in the table). The attrition indicator is a dummy which equals 1 if an individual leaves the sample in our data period 2006-2009. The attrition indicator is not significant for all subgroups.

C Price equations formal childcare

For non-users of formal childcare we have to simulate a price for childcare. We have information on the use of formal childcare in the Netherlands for the period 2006–2009. Here, a distinction is made between daycare (children 0–3 years of age) and out-of-school care (children 4–11 years of age).

Again, we estimate a quasi-fixed effects model for the prices of daycare and out-of-school care.[59] Here, we follow the same procedure as for the wage estimations and estimate the following price equation:

$$p_{it} = x'_{it}\beta + \omega_i + \bar{x_i}'\pi + \lambda_t invmills_{it} + \epsilon_{it} \tag{C.1}$$

where the individual specific effect consists of a random part, ω_i with $\sim IID(0, \sigma^2_\omega)$, and a part which is allowed to be correlated with regressors $\bar{x_i}'\pi$. Here, $\bar{x_i}$ is the average of age, which does not vary over time. Our dependent variable is the hourly real price.

We focus on households since childcare is consumed at the household level. As it turns out, characteristics of females are more important in predicting the use and gross price of childcare than characteristics of men. Hence, we only include females characteristics in the regressions.

Table C.1 shows estimation results for daycare and out-of-school care.[60] Estimation results show that year dummies are significantly increasing for daycare. However, time effects are less important in the price equation for out-of-school care. Households with higher educated women or younger women pay a higher price on average. We do not find evidence that selection bias (or attrition bias) is present.[61]

[59] We conduct a Hausman test in order to test whether fixed or random effects is appropriate. In all cases, the Hausman test favors the fixed effects model.

[60] Including a squared term for age, age splines, ethnicity, a dummy for age of the youngest child or a dummy for multiple children one at a time, leads to insignificant coefficients for each of these variables.

[61] However, note that for childcare prices the identification is solely off the functional form.

Table C.1: Price equation formal childcare

	Daycare	Out-of-school care
Year		
2007	0.074***	0.008
2008	0.149***	0.009
2009	0.183***	0.014
Higher educated woman[a]	0.036	-0.002
Age woman	-0.017***	-0.028***
Single parent	0.030**	-0.057**
Province[a]		
Groningen	0.027	0.063***
Friesland	0.153***	0.151***
Drenthe	0.156***	0.132***
Overijssel	0.121***	0.192***
Flevoland	-0.232***	-0.112***
Gelderland	0.116***	0.176***
Noord-Brabant	-0.035	-0.085**
Zeeland	-0.034***	0.010
Limburg	-0.019	0.118***
Urban area		
>150.000 habitants	0.095***	0.041***
Mundlak age average	0.012**	0.024***
Inverse Mills	0.062	-0.058
Constant	5.372***	5.709***
Observations	35,116	31,308
Number of households	14,830	13,193

*** p<0.01, ** p<0.05, * p<0.1. [a]Reference group: lower educated women in couples/single parents and living in the Western region of the Netherlands.

D Descriptive statistics

Table D.1: Descriptive statistics singles

	Singles without children		Single parents 0–3 yrs		Single parents 4–11 yrs	
	Mean	SD	Mean	SD	Mean	SD
Age	41.06	11.01	33.82	6.13	40.15	5.87
Male	0.58	0.49	0.04	0.20	0.08	0.27
Female	0.42	0.49	0.96	0.20	0.92	0.27
Native	0.79	0.41	0.54	0.50	0.65	0.48
Western immigrant	0.11	0.31	0.11	0.31	0.11	0.31
Non-Western immigrant	0.10	0.30	0.35	0.48	0.23	0.42
Lower educated	0.25	0.44	0.42	0.49	0.35	0.48
Middle educated	0.43	0.49	0.38	0.49	0.43	0.50
Higher educated	0.32	0.47	0.20	0.40	0.22	0.41
Large city	0.34	0.47	0.41	0.49	0.32	0.47
Small city	0.66	0.47	0.59	0.49	0.68	0.47
Hourly gross wage	16.76	7.79	15.63	7.27	16.23	8.00
Participation rate	0.89	0.32	0.57	0.50	0.70	0.46
Hours worked per week	35.61	7.87	28.80	8.50	28.53	8.41
Using formal childcare	0.00	0.00	0.51	0.50	0.27	0.44
Hours formal childcare per week	0.00	0.00	28.86	18.57	13.06	9.05
Observations	202,079		4,171		14,793	

	Single parents 12–17 yrs		Single parents 18+		Adult child	
	Mean	SD	Mean	SD	Mean	SD
Age	46.61	5.19	51.94	5.02	34.53	11.10
Male	0.18	0.38	0.27	0.45	0.72	0.45
Female	0.82	0.38	0.73	0.45	0.28	0.45
Native	0.69	0.46	0.69	0.46	0.80	0.40
Western immigrant	0.12	0.32	0.13	0.33	0.10	0.30
Non-Western immigrant	0.19	0.39	0.19	0.39	0.10	0.31
Lower educated	0.34	0.47	0.35	0.48	0.47	0.50
Middle educated	0.41	0.49	0.40	0.49	0.41	0.49
Higher educated	0.24	0.43	0.25	0.43	0.12	0.33
Large city	0.30	0.46	0.31	0.46	0.21	0.40
Small city	0.70	0.46	0.69	0.46	0.79	0.40
Hourly gross wage	16.47	7.74	17.11	8.38	12.78	5.32
Participation rate	0.81	0.39	0.82	0.39	1.00	0.06
Hours worked per week	31.18	9.03	33.01	8.94	35.70	8.69
Using formal childcare	0.00	0.00	0.00	0.00	0.00	0.00
Hours formal childcare per week	0.00	0.00	0.00	0.00	0.00	0.00
Observations	20,767		9,171		12,544	

Table D.2: Descriptive statistics couples

	Couples without children				Couples with children 0–3 yrs				Couples with children 4–11 yrs			
	Men		Women		Men		Women		Men		Women	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	46.70	11.62	44.61	11.98	37.27	4.84	34.60	4.26	43.43	4.93	41.01	4.45
Native	0.87	0.33	0.86	0.35	0.85	0.36	0.84	0.37	0.85	0.36	0.84	0.37
Western immigrant	0.09	0.29	0.10	0.30	0.08	0.27	0.08	0.27	0.08	0.27	0.09	0.29
Non-Western immigrant	0.03	0.18	0.04	0.20	0.08	0.27	0.08	0.27	0.07	0.25	0.07	0.25
Lower educated	0.24	0.43	0.30	0.46	0.20	0.40	0.15	0.36	0.24	0.42	0.23	0.42
Middle educated	0.43	0.50	0.42	0.49	0.43	0.50	0.49	0.50	0.43	0.50	0.52	0.50
Higher educated	0.33	0.47	0.28	0.45	0.37	0.48	0.36	0.48	0.33	0.47	0.25	0.43
Large city	0.18	0.38	0.18	0.38	0.17	0.37	0.17	0.37	0.15	0.35	0.15	0.35
Small city	0.82	0.38	0.82	0.38	0.83	0.37	0.83	0.37	0.85	0.35	0.85	0.35
Hourly gross wage	20.39	10.53	15.75	6.76	20.43	10.15	16.45	7.01	22.28	11.45	16.27	7.28
Participation rate	0.96	0.19	0.79	0.41	0.97	0.18	0.81	0.40	0.96	0.20	0.76	0.43
Hours worked per week	38.06	6.19	29.30	9.81	38.68	5.20	22.80	8.05	38.78	5.45	21.36	8.55
Using formal childcare	0.00	0.00	0.00	0.00	0.46	0.50	0.46	0.50	0.13	0.34	0.13	0.34
Hours formal childcare per week	0.00	0.00	0.00	0.00	20.22	11.41	20.22	11.41	8.96	6.72	8.96	6.72
Observations	134,687		134,687		13,849		13,849		14,723		14,723	

	Couples with children 12–17 yrs				Couples with children 18+ yrs			
	Men		Women		Men		Women	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	48.82	4.60	46.56	4.24	54.19	4.34	52.26	4.35
Native	0.86	0.35	0.85	0.36	0.87	0.34	0.86	0.35
Western immigrant	0.08	0.28	0.09	0.29	0.09	0.29	0.09	0.29
Non-Western immigrant	0.06	0.23	0.06	0.23	0.05	0.21	0.05	0.22
Lower educated	0.24	0.43	0.29	0.45	0.27	0.44	0.40	0.49
Middle educated	0.42	0.49	0.47	0.50	0.40	0.49	0.41	0.49
Higher educated	0.34	0.47	0.24	0.43	0.33	0.47	0.19	0.39
Large city	0.15	0.35	0.15	0.35	0.14	0.35	0.14	0.35
Small city	0.85	0.35	0.85	0.35	0.86	0.35	0.86	0.35
Hourly gross wage	22.93	12.11	15.91	7.29	22.85	12.26	15.25	6.27
Participation rate	0.96	0.20	0.79	0.41	0.96	0.20	0.71	0.45
Hours worked per week	38.45	6.10	23.88	9.32	38.01	6.21	24.08	9.58
Using formal childcare	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hours formal childcare per week	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	58,988		58,988		10,892		10,892	

	Couples without children				Couples with children			
	Men		Women		Men		Women	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	50.29	10.42	52.06	9.92	43.35	6.96	41.00	7.36
Male	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Female	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
Native	0.84	0.36	0.86	0.35	0.80	0.40	0.80	0.40
Western immigrant	0.11	0.31	0.10	0.30	0.10	0.30	0.09	0.28
Non-Western immigrant	0.04	0.21	0.04	0.19	0.10	0.30	0.12	0.32
Lower educated	0.28	0.45	0.44	0.50	0.23	0.42	0.26	0.44
Middle educated	0.39	0.49	0.36	0.48	0.39	0.49	0.47	0.50
Higher educated	0.33	0.47	0.19	0.40	0.38	0.49	0.27	0.44
Large city	0.18	0.39	0.16	0.36	0.19	0.39	0.17	0.38
Small city	0.82	0.39	0.84	0.36	0.81	0.39	0.83	0.38
Hourly gross wage	21.22	11.92	15.86	7.06	22.60	12.89	16.64	8.15
Participation rate	0.95	0.21	0.58	0.49	0.95	0.21	0.67	0.47
Hours worked per week	37.38	6.74	27.25	10.27	38.25	6.77	23.19	9.53
Using formal childcare	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hours formal childcare per week	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	16,486		32,962		24,070		37,034	

E Preferences and fit of structural models

For some household types the full translog specification resulted in a significant share (>5%) of households with negative marginal utility of income in the observed choices. This is not consistent with utility maximisation and drives down the labour supply elasticities to implausible values. For these household types we dropped the interaction terms between income and leisure, which resulted in a low share of households with negative marginal utility of income (<5%). For some households we also obtained an 'inverted' pattern for the marginal utility of income, with a negative (log) linear term and a positive (log) quadratic term. This results in implausible (positive) income effects, and for these households we dropped the quadratic term in income. Finally, for certain other household types, the translog specification was still not flexible enough. In particular, in some cases we do not capture the distribution of hours worked at the top very well, and we introduce a third-order term for (log) leisure, which then improves the fit at the top.

Table E.1: Estimated preferences singles

Parameters	Singles	Single parents Youngest child 0-3	Single parents Youngest child 4-11	Single parents Youngest child 12-17	Single parents Youngest child 18+
Income	0.140***	5.412***	3.174***	1.507***	1.230***
Income^2	0.237***				
Leisure	46.058***	-45.840***	-64.770***	-5.172***	27.033***
X (age-38)/10	2.238***	-2.303***	-2.327***	-1.783***	-1.847***
X (age-38)^2/100	0.842***	-0.199***	2.888***	1.964***	1.909***
Leisure^2	469.715***	-200.500***	-226.800***	125.441***	336.933***
Leisure^3	1009.912***			445.761***	816.440***
St. dev. Leisure	2.203***	14.910***	0.908***	2.725***	1.482***
Fixed costs of work	0.158***	-3.146***	-2.365***	-1.272***	-0.194***
X 1(low educated)	-0.820***	-1.964***	-2.494***	-1.362***	-1.261***
X 1(medium educated)	-0.071***	-0.192***	-0.607***	-0.305***	-0.379***
X 1(non-Western allochtonous)	-1.335***	-1.693***	-2.803***	-1.217***	-1.105***
X 1(Western allochtonous)	-0.626***	-0.873***	-1.221***	-0.710***	-0.687***
X 1(>=150,000 inhabitants)		-0.557***	-0.630***		
Hours of formal childcare		-0.329***	-1.231***		
X 1(non-Western allochtonous)		0.978***	1.423***		
X 1(Western allochtonous)		0.501***	0.217***		
X 1(>=150,000 inhabitants)		0.320***	0.302***		
Hours of formal childcare^2		-0.062***	-0.202***		
St.dev. Hours childcare		-1.273***	0.908***		
Fixed costs of childcare		0.163***	-1.382***		
X 1(low educated)		-2.628***	-1.340***		
X 1(medium educated)		-1.036***	-0.614***		
X 1(non-Western allochtonous)		-2.354***	-1.262***		
X 1(Western allochtonous)		-0.793***	-0.198***		
Income X leisure					
Income X hours of formal childcare		-0.325***	19.140***		
Leisure X hours of formal childcare		-7.126***	-7.126***		
Observations	202,079	4,168	14,789	20,767	9,171

*** p<0.01, ** p<0.05, * p<0.1.

Table E.2: Estimated preferences couples

Parameters	Couples w/o children	Couples Young. child 0-3	Couples Young. child 4-11	Couples Youngest child 12-17	Couples Youngest child 18+
Income	2.925***	10.390***	3.920***	0.889***	0.636***
Income^2	0.179***	-0.511***	1.338***	1.585***	1.513***
Leisure man	75.790***	-72.630***	-87.450***	79.187***	86.479***
X (age man-38)/10	2.071***	3.995***	2.771***	-0.197***	-1.107***
X (age man-38)^2/100	0.441***	-0.157***	1.261***	2.364***	1.609***
Leisure man^2	671.165***	-122.500***	-118.300***	827.839***	824.354***
Leisure man^3	1298.248***			1634.206***	1575.481***
St. dev. leisure man		9.877***	9.545***	8.329***	0.639***
Leisure woman	-0.469***	-21.170***	-19.170***	-44.378***	-31.408***
X (age woman-38)/10	5.138***	3.382***	0.614***	1.550***	2.273***
X (age woman-38)^2/100	1.356***	2.843***	0.559***	0.629***	1.505***
Leisure woman^2	107.909***	-159.900***	-109.200***	-218.239***	-111.450***
Leisure woman^3	410.896***			-254.075***	-33.661***
St. dev. leisure woman		-3.199***	-1.260***		0.132***
Fixed costs of work man	0.661***	-9.643***	-9.358***	1.120***	1.265***
X 1(low educated man)	0.317***	1.411***	0.629***	0.237***	0.196***
X 1(medium educated man)	0.413***	1.411***	1.073***	0.712***	0.280***
X 1(non-West. allocht. man)	-1.706***	-1.177***	-1.935***	-2.227***	-1.324***
X 1(Western allocht. man)	-1.098***	-1.609***	-1.563***	-1.237***	-0.776***
Fixed costs of work woman	-1.062***	-2.765***	-1.690***	-2.257***	-1.905***
X 1(low educated woman)	-1.052***	-0.221***	-0.579***	-1.084***	-1.152***
X 1(medium educated woman)	-0.318***	0.309***	-0.097***	-0.399***	-0.430***
X 1(non-West. allocht. woman)	-1.196***	-1.230***	-0.540***	-0.494***	-0.156***
X 1(Western allocht. woman)	-0.357***	-0.603***	-0.218***	-0.223***	-0.098***
Hours of formal childcare		-2.016***	-1.100***		
X 1(non-West. allocht. man)		-0.0162***	0.108***		
X 1(Western allocht. man)		0.400***	0.178***		
X 1(non-West. allocht. woman)		0.903***	0.603***		
X 1(Western allocht. woman)		0.352***	0.194***		
X 1(>=150,000 inhabitants)		0.413***	0.186***		
Hours of formal childcare^2		-0.101***	-0.687***		
St. dev. hours childcare		0.010*	0.025**		
Fixed costs of childcare		0.141***	-2.304***		
X 1(low educated man)		-0.418***	-0.275***		
X 1(medium educated man)		-0.157***	-0.166***		
X 1(non-West. allocht. man)		-0.551***	-0.286***		
X 1(Western allocht. man)		-0.689***	-0.213***		
X 1(low educated woman)		-1.041***	-1.359***		
X 1(medium educated woman)		-0.588***	-0.576***		
X 1(non-West. allocht. woman)		-1.534***	-0.394***		
X 1(Western allocht. woman)		-0.622***	-0.153***		
X 1(>=150,000 inhabitants)		-0.315***	0.077***		
Income X leisure man	5.867***	4.432***	11.750***	13.666***	15.891***
Income X leisure woman	-2.071***	-3.098***	2.576***	3.694***	3.191***
Leisure man X leisure woman	12.447***	-21.340***	2.363***	19.047***	31.912***
Income X hours of formal childcare		0.616***	0.522***		
Leisure man X hrs of formal childcare		0.076***	-0.582***		
Leisure woman X hrs of formal childcare		-6.581***	-7.036***		
Observations	134,687	13,849	14,723	58,988	10,892

*** p<0.01, ** p<0.05, * p<0.1.

Table E.3: Estimated preferences couples with only one flexible partner and adult children

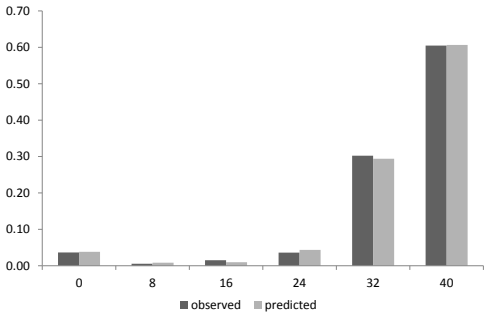
Parameters	Couples w/o children man flexible	Couples w/o children woman flex.	Couples w/o children man flexible	Couples w/o children woman flex.	Adult child
Income	1.109***	1.411***	3.827***	1.170***	3.084***
Income^2		0.236***	-0.520***	0.630***	-2.338***
Leisure	22.875***	-14.856***	77.732***	-44.399***	62.083***
X (age-38)/10	3.451***	3.423***	1.469***	-1.751***	-0.523***
X (age-38)^2/100	0.713***	1.800***	0.838***	1.342***	2.114***
Leisure^2	507.038***	40.070***	628.002***	-228.552***	351.100***
Leisure^3	1193.080***	251.237***	1211.015***	-304.791***	732.437***
Standard dev. Leisure	17.730***		0.253***	0.129***	
Fixed costs of work	0.327***	-1.735***	0.681***	-2.351***	5.802***
X 1(low educated)	0.503***	-1.209***	-0.107***	-1.336***	-0.944***
X 1(medium educated)	-0.236***	-0.447***	-0.030***	-0.638***	-0.605***
X 1(non-West. allocht.)	-1.869***	-1.307***	-1.035***	0.069***	-2.096***
X 1(Western allocht.)	-1.738***	-0.127***	-0.716***	-0.512***	-0.622***
X 1(young. child 4-11)			0.060***	-1.108***	
X 1(young. child 12-17)			0.330***	-0.315***	
X 1(young. child 18+)			0.551***	-0.111***	
Income X leisure				3.241***	-21.860***
Observations	16,486	32,962	24,070	37,034	12,544

*** p<0.01, ** p<0.05, * p<0.1.

F Fit hours distribution

Figure F.1: Couples without children

(a) Men



(b) Women

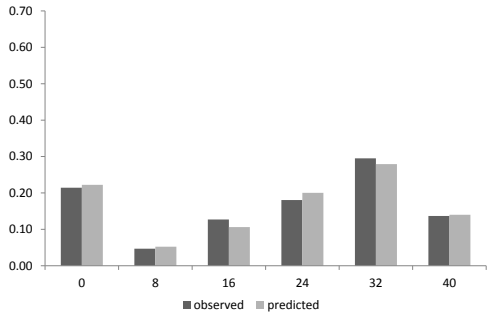
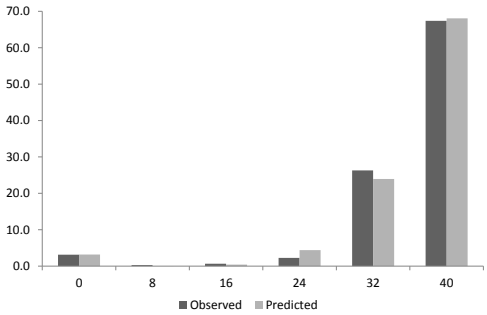


Figure F.2: Couples with youngest child 0–3

(a) Men



(b) Women

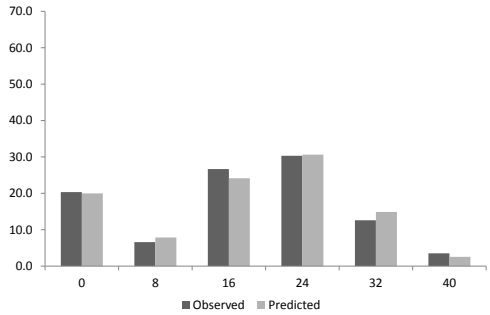
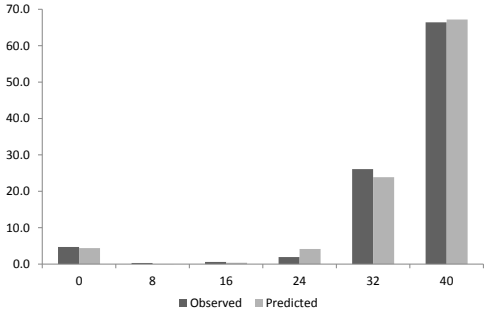


Figure F.3: Couples with youngest child 4–11

(a) Men



(b) Women

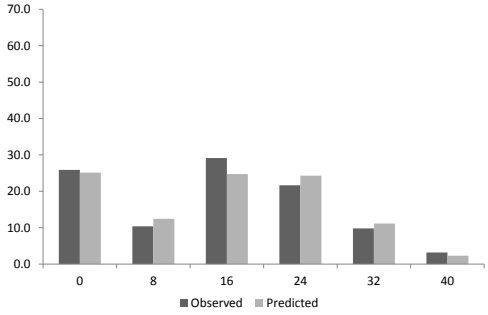
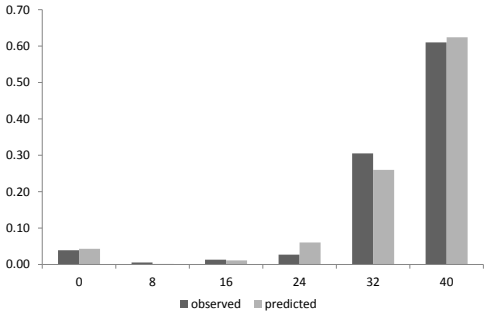


Figure F.4: Couples with youngest child 12–17

(a) Men



(b) Women

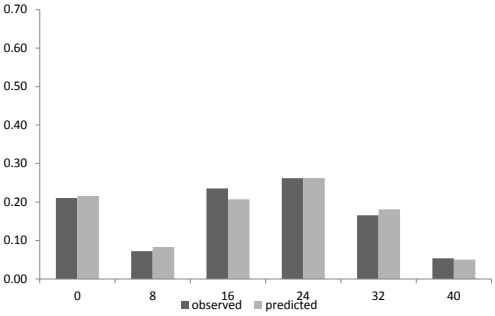
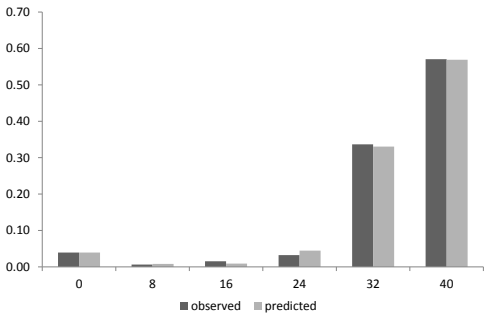


Figure F.5: Couples with youngest child 18+

(a) Men



(b) Women

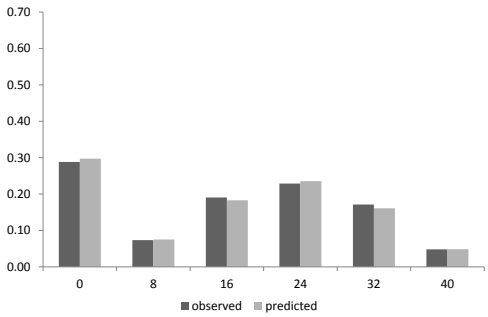
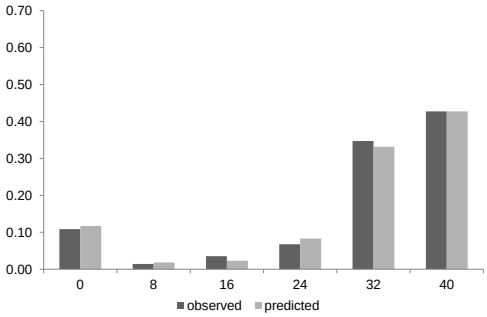


Figure F.6: Singles and single parents youngest child 0–3

(a) Singles



(b) Single parents young. child 0–3

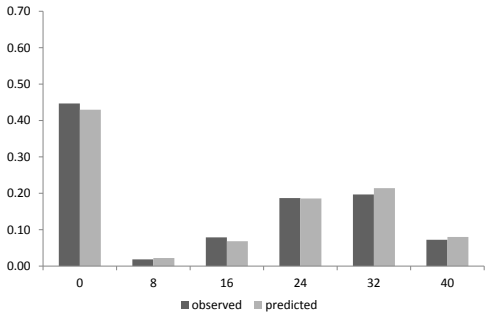
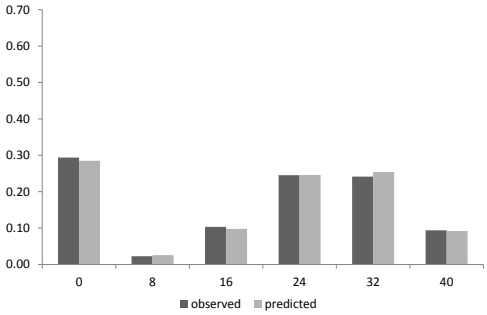


Figure F.7: Single parents youngest child 4–11 and 12–17

(a) Single parents young. child 4–11



(b) Single parents young. child 12–17

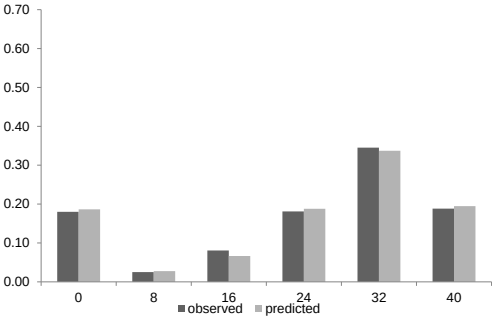
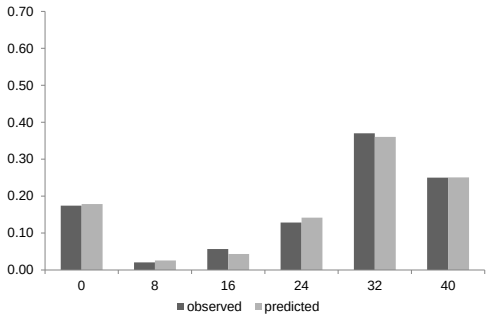


Figure F.8: Single parents youngest child 18+ and adult children

(a) Single parents young. child 18+



(b) Adult child living at home

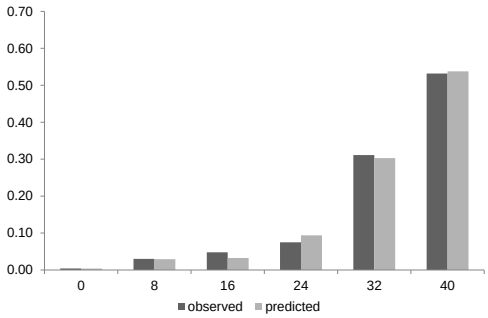
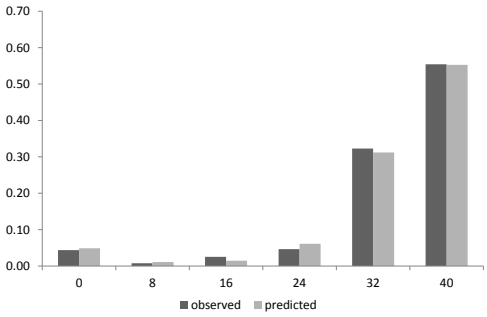


Figure F.9: Couples without children: inflexible partner

(a) Men



(b) Women

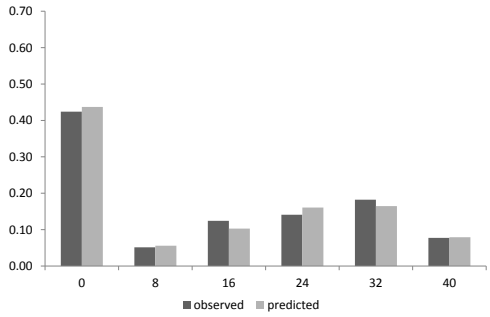
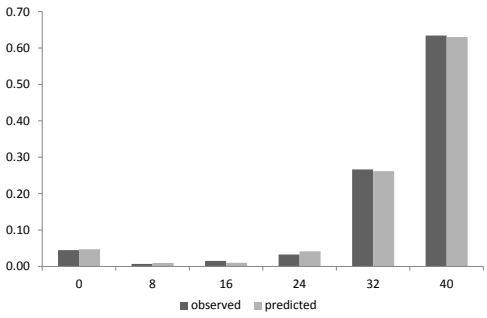
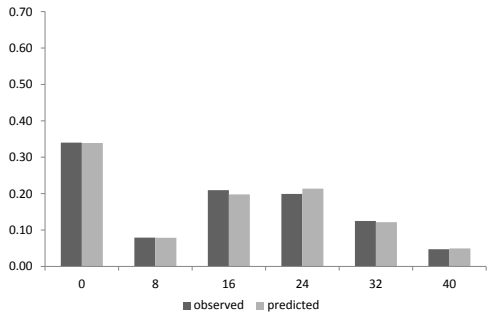


Figure F.10: Couples with children: inflexible partner

(a) Men



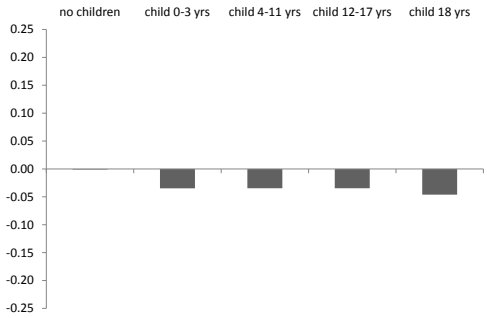
(b) Women



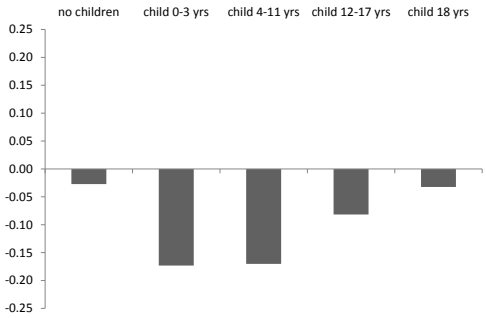
G Cross-elasticities in couples

Figure G.1: Cross elasticities in households with two flexible persons

(a) Men



(b) Women



H Preparing the Income Panel for policy simulations

The first step is to construct all the relevant variables for the labour supply module using the data from the Income Panel. In the Income Panel we do not have education. We impute education in the Income Panel using demographic characteristics observed in both the Income Panel (for the year 2010) and the Labour Market Panel (for the year 2009). For the employed we also include hourly wages as an explanatory variable next to the demographic variables. We classify education as follows (using the Dutch abbreviations): i) lower educated = BO and VMBO, ii) middle educated = MBO, HAVO and VWO, iii) higher educated = HBO and WO. We estimate ordered logit models separately for each household type (both sexes in couples), and separately for workers and non-workers. We then predict the level of education for each individual using data from the Labour Market Panel. The resulting education distribution per household type is very similar to the Labour Market Panel, see Table H.2.[62] Next, for workers we observe the hourly gross wage in the Income Panel. For non-workers we need to impute (potential) hourly gross wages. We use the estimated wage equations of the Labour Market Panel[63], given in Appendix B to impute wages for non-workers in the Income Panel.[64] Finally, for users of formal childcare we observe the hourly gross price. For non-users we impute the (potential) hourly gross price. We use the estimated price equations of the Labour Market Panel given in Appendix C to impute childcare prices for non-users in the Income Panel.

We also want to include the responses of individuals with unemployment benefits. Here we follow the method by Ericson and Flood (2012). First, we need to calculate the difference in income from UI and wage earnings. This is typically not observed in the data. We either observe wages or we observe the level of UI. For each individual we need to simulate income in the alternative situation. For individuals with UI, we impute an hourly wage and use this wage to predict gross labour income when working 40 hours per week for the whole year. For workers we predict the gross UI income when unemployed for 40 hours per week (taking into account the maximum income over which unemployment insurance applies). The replacement rate is then defined as disposable household income when the individual is on unemployment benefits over disposable household income when the individuals is employed, taking the income and employment status of the partner as given. Next, we estimate a logit model for the probability of being in UI, using data for the period 2006–2009 (in line with

[62] There are minor differences, which are inevitable given the (apparent) sampling variation in the characteristics (both datasets are representative samples of the Dutch population using the respective weights supplied by Statistics Netherlands).

[63] Applying the appropriate correction for wage inflation, moving from 2009 to 2010.

[64] Using the imputed education level. This guarantees that education and wages in the Income Panel keep the correlation in the data in the Labour Market Panel.

the rest of the structural model). The individual and household characteristics have the expected sign. However, the replacement rate always has the 'wrong' sign, where a higher replacement rate lowers the probability of being in UI. This is probably due to insufficient exogenous variation in replacement rates for UI in our data. However, we can still use the methodology of Ericson and Flood (2012) to incorporate UI responses in the model, calibrating the response to changes in the replacement rate on the findings of empirical studies. After calculating the replacement rate in UI for all individuals, we include it in the logit model to calibrate the elasticity of the probability of being in UI with respect to the replacement rate. We follow the CPB study *Kansrijk Arbeidsmarktbeleid* (CPB, 2015, p. 83) and use 0.5 for the elasticity of the probability of being in UI with respect to the benefit level. Then, we calibrate the model such that it generates the desired average elasticity over all groups. Taking draws from the error term distribution we can predict whether or not an individual receives UI or not (0 or 1). Individuals on UI do not enter the discrete choice model. Individuals who do not receive UI do enter the discrete choice model. When the replacement rate changes, the status of an individual of being in or out of UI may change.

Finally, we want to simulate policy changes in a recent year, rather than over the estimation period (2006–2009). We use 'static ageing' to generate a dataset for 2015 that is consistent with the demographic projection from Statistics Netherlands. Furthermore, we update all the price and income variables to those of 2015 (keeping the real values, relevant for utility, in 2006 prices). Finally, we update the parameters of the tax-benefit system to 2015.

Table H.1: Individual and household characteristics: IPO 2010 versus LMP 2009

	Total	Men in couples w/o child	Women in couples w/o child	Men in couples with child	Women in couples with child	Singles	Single parents	Adult child
Labour Market Panel								
Female	0.513	0.000	1.000	0.000	1.000	0.417	0.842	0.342
Age	41.285	43.580	45.041	43.665	41.670	39.706	43.660	24.703
Native	0.811	0.850	0.850	0.820	0.807	0.776	0.662	0.813
Non-Western immigrant	0.096	0.056	0.052	0.097	0.103	0.115	0.226	0.111
Western immigrant	0.093	0.093	0.099	0.083	0.090	0.109	0.112	0.076
0-49 inhabitants[b]	0.494	0.481	0.514	0.538	0.541	0.336	0.375	0.561
50-149 inhabitants	0.294	0.297	0.293	0.292	0.288	0.307	0.319	0.275
>149 inhabitants	0.213	0.222	0.193	0.171	0.171	0.357	0.307	0.164
Northern region	0.101	0.103	0.111	0.099	0.103	0.098	0.093	0.087
Eastern region	0.213	0.210	0.211	0.227	0.227	0.175	0.179	0.226
Western region	0.469	0.461	0.448	0.452	0.452	0.542	0.537	0.455
Southern region	0.217	0.225	0.230	0.222	0.219	0.184	0.192	0.232
Gross hourly wage[a]	18.596	20.998	16.523	22.863	16.892	17.611	17.184	11.920
Income Panel								
Female	0.504	0.000	1.000	0.000	1.000	0.422	0.846	0.321
Age	41.098	44.397	44.184	43.626	41.234	39.903	43.302	26.062
Native	0.788	0.840	0.830	0.816	0.799	0.709	0.648	0.793
Non-Western immigrant	0.111	0.065	0.061	0.103	0.110	0.150	0.239	0.131
Western immigrant	0.101	0.095	0.108	0.081	0.091	0.141	0.114	0.076
0-49 inhabitants[b]	0.467	0.468	0.485	0.516	0.522	0.317	0.369	0.530
50-149 inhabitants	0.296	0.308	0.300	0.295	0.291	0.298	0.318	0.273
>149 inhabitants	0.237	0.224	0.215	0.189	0.187	0.385	0.313	0.197
Northern region	0.100	0.104	0.107	0.101	0.103	0.095	0.095	0.085
Eastern region	0.210	0.210	0.210	0.225	0.223	0.177	0.183	0.218
Western region	0.479	0.459	0.453	0.458	0.460	0.551	0.537	0.473
Southern region	0.211	0.227	0.230	0.216	0.215	0.177	0.186	0.225
Gross hourly wage[a]	17.736	20.002	15.968	21.662	16.594	16.560	16.696	11.520

[a]In 2009 prices. [b]Degree of urbanisation (x1000 inhabitants).

Table H.2: Education: predicted IPO 2010 versus observed LMP 2009[a]

	Total	Men in couples w/o child	Women in couples w/o child	Men in couples with child	Women in couples with child	Singles	Single parents	Adult child
Labour Market Panel								
Lower educated	0.307	0.254	0.359	0.240	0.275	0.283	0.360	0.597
Medium educated	0.427	0.437	0.401	0.431	0.467	0.425	0.421	0.337
Higher educated	0.265	0.309	0.240	0.329	0.258	0.292	0.218	0.065
Income Panel								
Lower educated	0.318	0.274	0.356	0.251	0.268	0.312	0.369	0.606
Medium educated	0.427	0.439	0.413	0.436	0.464	0.420	0.422	0.330
Higher educated	0.255	0.288	0.232	0.313	0.267	0.268	0.209	0.064

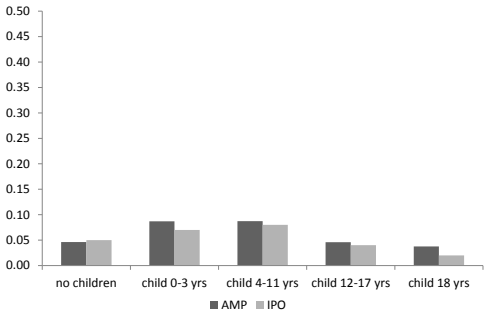
[a]Education is classified as follows (using the Dutch abbreviations): i) lower educated = BO and VMBO, ii) middle educated = MBO, HAVO and VWO, iii) higher educated = HBO and WO.

I Elasticities LMP versus IPO

We compare the behavioural responses using the enriched and uprated Income Panel for 2015 with the behavioural responses using data from the Labour Market Panel for 2009. We see that the labour supply elasticities are very similar. Labour supply elasticities are somewhat lower for women with young children because their participation rate is higher in 2015, which is partly the result of a lower participation tax in 2015 than in 2009.

Figure I.1: Households with two flexible persons

(a) Men



(b) Women

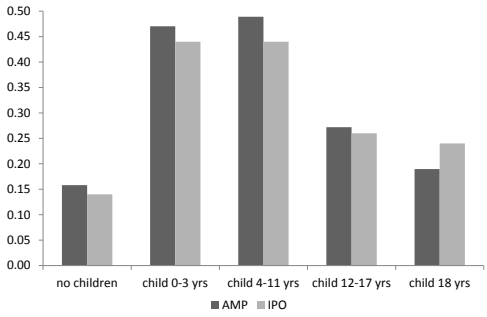
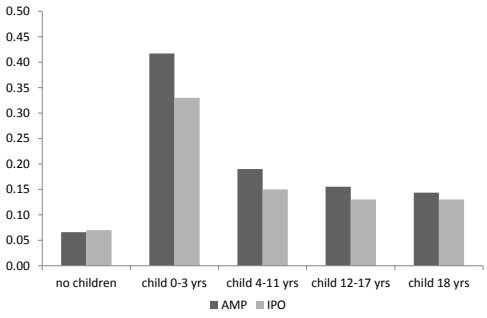


Figure I.2: Households with one flexible person, and adult children

(a) Singles and single parents



(b) Individuals with an inflexible partner, and adult children living at home