

DISCUSSION PAPER SERIES

No. 9275

TAXATION AND THE LONG RUN ALLOCATION OF LABOR: THEORY AND DANISH EVIDENCE

Claus Thustrup Kreiner, Jakob Roland
Munch and Hans Jørgen Whitta-Jacobson

PUBLIC POLICY



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP9275.asp

TAXATION AND THE LONG RUN ALLOCATION OF LABOR: THEORY AND DANISH EVIDENCE

Claus Thustrup Kreiner, University of Copenhagen, CESifo and CEPR
Jakob Roland Munch, University of Copenhagen and IZA
Hans Jørgen Whitta-Jacobson, University of Copenhagen

Discussion Paper No. 9275
January 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **PUBLIC POLICY**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Claus Thustrup Kreiner, Jakob Roland Munch and Hans Jørgen Whitta-Jacobson

ABSTRACT

Taxation and the Long Run Allocation of Labor: Theory and Danish Evidence*

Inspired by Hayek (1945), we study the distortionary effects of taxation on labor mobility and the long run allocation of labor across different profitable opportunities. These effects are not well detected by the methods applied in the large public finance literature estimating the elasticity of taxable income and quantifying the welfare loss from taxation. Our analysis builds on a standard search theoretic framework where workers are continually seeking better paid jobs, but are also fired from time to time because of economic development and productivity shocks. We incorporate non-linear taxation into this setting and estimate the structural parameters of the model using employer-employee register based data for the full Danish population of workers and workplaces for the years 2004-2006. Our results indicate that along the intensive margin the Danish taxation generates an overall efficiency loss corresponding to a 12 percent reduction in GDP. It is possible to reap 4/5 of this potential efficiency gain by going from a high-tax Scandinavian system to a level of taxation in line with low-tax OECD countries such as the United States. The tax-responsiveness of labor mobility and allocation corresponds to an elasticity of taxable income with respect to the net-of-tax rate in the range 0.15-0.3.

JEL Classification: H21

Keywords: elasticity of taxable income, labor mobility and tax distortions

Claus Thustrup Kreiner
Department of Economics
University of Copenhagen
Øster Farimagsgade 5, building 26
DK-1353 Copenhagen K
DENMARK

Jakob Roland Munch
Department of Economics
University of Copenhagen
Øster Farimagsgade 5, building 26
DK-1353 Copenhagen K
DENMARK

Email: ctk@econ.ku.dk

Email:
jakob.roland.munch@econ.ku.dk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154849

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=143981

Hans Jørgen Whitta-Jacobson
Department of Economics
University of Copenhagen
Øster Farimagsgade 5, building 26
DK-1353 Copenhagen K
DENMARK

Email: jacob@econ.ku.dk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=152389

*We thank participants at the Conference on the Economics of the Nordic Model in Oslo for helpful comments and suggestions. Financial support from the Economic Policy Research Network (EPRN) is gratefully acknowledged.

Submitted 30 December 2012

1 Introduction

Hayek (1945) argues convincingly that the true merit of the capitalist market economy is its capability of continually responding to the ongoing changes in needs and possibilities by reallocating resources towards their best use.¹ Prices convey information about productivity and scarcity of resources, and owners of resources react on price signals to obtain the highest possible returns, thereby collectively acting as an effective allocation device. According to this view, taxation distorts the functioning of the market economy by creating a wedge between the private return and the social return to a reallocation of resources, leaving socially desirable opportunities unexploited as a result. This type of tax distortion should be relevant for the labor market since employees change jobs many times during their career and economic incentives are a key determinant behind job changing decisions (Topel and Ward, 1992).

This paper studies the impact of taxation on the mobility and allocation of labor, and quantifies the efficiency loss from misallocation of labor caused by taxation.² As will be explained below, labor mobility responses are fundamentally different from the hours-of-work responses of the basic labor supply model, which underlies the modern reduced-form empirical identification methods to measure behavioral responses to taxation. These methods are not well suited to detect mobility behavior, and therefore we apply a structural approach. Our analysis builds on a standard search theoretic framework, along the lines of Burdett (1978) and Christensen et al. (2005), where workers continually seek better paid jobs, but are also fired from time to time because of economic development and productivity shocks. We incorporate non-linear taxation into this setting and estimate the structural parameters of the model using employer-employee register based data for the full Danish population of workers and workplaces for the years 2004-2006. The estimated model is then used to examine the impact of different changes in the tax system, thereby characterizing the distortionary effects of taxation on the allocation of labor.

The overall tax distortion is identified by an experiment where all taxation along the intensive margin is replaced by lump sum taxation. For Denmark, with marginal tax rates ranging from 59 to 72 percent, the result is an increase in GDP of 1/4, a welfare gain equal to 12 percent of GDP and a degree of self-financing of 13 percent.³ This is a sizable, although not huge, efficiency loss from taxation. Interestingly, it is possible to reap a very large part of the potential efficiency gain

¹The seminal article by Hayek was selected in 2011 to be one of the twenty most influential articles published by the American Economic Review in the entire history of the review.

²Our focus is on measuring behavioral responses to taxation and quantifying the efficiency loss from taxation. This is in line with the literature estimating the elasticity of taxable income. We do not attempt to quantify the trade-off between equality and efficiency (see Immervoll et al, 2007), but a normative tax analysis should, of course, balance the efficiency loss from taxation against the distributional consequences.

³The degree of self-financing measures the share of the mechanical tax revenue loss recouped by behavioral responses. For small tax changes, the degree of self-financing is approximately equal to the marginal excess burden of taxation measured as the welfare effect in proportion to the mechanical revenue change.

by going “half the way” and replace the current taxation with a flat tax of 30 percent. This shift from a Scandinavian tax system with high marginal tax rates to a level of taxation in line with low-tax OECD countries such as the United States increases GDP by 20 percent, yields a welfare gain measured in proportion to GDP of 10 percent, and has a degree of self-financing of about 40 percent. The large welfare effect from going only “half the way” reflects the well-known result in the economics of taxation that the marginal excess burden is growing in the level of taxation. The structural approach enables us to derive this relationship between the marginal excess burden and the level of taxation, which turns out to be strongly convex with a Laffer rate around 2/3. Finally, and most importantly, we compute the marginal excess burden at the initial tax system by looking at a small reduction in all marginal tax rates.⁴ The degree of self-financing and the marginal excess burden are in this case equal to 87 percent of the (mechanical) loss in tax revenue, which translates into an elasticity of labor market income (taxable income) with respect to the net-of-tax rate of 0.3. A number of sensitivity tests suggest that the estimate of the elasticity of taxable income due to mobility responses lies in the interval 0.15 to 0.3.

Our results are related to the large public finance literature, recently surveyed by Saez et al. (2012), estimating the elasticity of taxable income with respect to the net-of-tax rate using tax return data. The ultimate goal of this literature is to estimate the ‘true’ underlying structural elasticity relevant for quantifying the efficiency losses from taxation. In two influential papers, Feldstein (1995, 1999) argued that all effects of the tax system on the reported income of the tax payers represent welfare effects, and therefore the elasticity of taxable income is a sufficient statistic for normative analysis. However, the inclusion of taxable income responses due to tax avoidance and evasion behavior relies on strong assumptions (Chetty, 2009; Slemrod and Kopczuk, 2002), and recent empirical studies have increasingly focused on more narrow income concepts, mainly labor income as we also do, in order to isolate real responses (e.g. Chetty et al., 2011a; Kleven and Schultz, 2012).

Most of the literature has exploited variation in tax rates generated by tax reforms to identify the elasticity using difference-in-difference type methodologies where individual post-reform income levels are compared to pre-reform levels. Saez et al. (2012) conclude that an elasticity of taxable income of 0.25 seems to be a reasonable estimate from the existing studies and the recent empirical evidence for Denmark in Kleven and Schultz (2012), based on a series of tax reforms and a very rich administrative data set covering the entire Danish population, points to an elasticity of labor income around 0.1.

Other recent studies have exploited bunching at the kink points of the tax system to identify the elasticity of taxable income (Saez, 2010; Chetty et al., 2011a, Bastani and Selin, 2012) and

⁴More specifically, we consider a one percent increase in the net-of-tax rates.

have found very small elasticities. Using Danish data, Chetty et al. (2011a) estimate an elasticity below 0.02. However, an important conclusion from this study is that the elasticity is larger when measured at larger kinks in the tax schedule. This strongly indicates that labor supply decision making is subject to short term frictions and adjustment costs, implying that the micro econometric identification methods in general underestimate the structural elasticity. After accounting for small frictions, corresponding to less than one percent of earnings, Chetty et al. (2011a,b) and Chetty (2012) conclude that the small elasticities obtained in the micro labor supply literature correspond to a structural intensive margin elasticity of around 0.3, more in line with the elasticity found in the macro literature based on cross-country evidence.

Both the difference-in-difference type methods and the bunching method are unlikely to identify the labor allocation effects of taxation studied here. To see why, consider a change of the tax system where first all marginal tax rates are identical and then the marginal tax rates for income levels above a certain threshold are raised to a new, higher level, thus introducing tax progressivity with a kink-point in the tax schedule. According to a standard, hours-of-work labor supply model with heterogeneity in abilities, the behavior of workers with income below the threshold will be unaffected, while all workers with income above the threshold will reduce labor supply and income. Hence, it is possible to separate workers into distinct control and treatment groups depending on income levels before the tax change, and the tax increase makes workers pile up at the income level of the kink-point, thereby creating a discontinuity in the density function. This makes it possible to identify the underlying structural labor supply elasticity by measuring changes in income of the treatment group relative to the control group or by exploiting the bunching of individuals at the kink-point. In the search framework, both workers with income above the threshold and below it will reduce search effort. The reason that also workers below the threshold reduce effort is that their expected after-tax income gain from searching and climbing up the income ladder is reduced. Therefore, they do not constitute a well-defined control group and the kink in the tax schedule does not create an excess mass of individuals at the income threshold.⁵

To conclude, the otherwise strong empirical identification strategies normally used to identify income responses to taxation and the elasticity of taxable income are not well-suited to detect the income responses due to job mobility effects, and our estimates of the elasticity of taxable income and of the tax distortions of labor allocation, based on a structural approach, indicate that these job mobility effects are of a substantial magnitude.

A few studies have analyzed the impact of taxation on job mobility decisions by looking at the mobility of people across states (Feldstein and Wrobel, 1998) and across countries (Kleven et

⁵Another reason why the difference-in-difference types techniques are unlikely to detect the effects on search behavior is that job shifts and reallocation of labor are time-consuming processes, implying that it takes relatively long time for the economy to reach its new long run distribution of earnings after a tax reform.

al., 2011). The second study examines the impact of changes in the preferential tax regime for foreign researchers and high-income foreigners in Denmark and finds large effects on the number of highly paid foreigners. These types of studies are not comparable to our analysis of the impact of taxes on worker mobility and job allocation within a country, but they do show that taxation is important for the decision to shift job, as also found in our analysis. A study by Gentry and Hubbard (2004) looks at the impact of taxes on within-country job mobility. They show, using a reduced-form empirical analysis of the relationship between job turnover and tax rates in the US (together with a number of other co-variates), that job shifting is negatively correlated with tax rates and tax progression.

The paper proceeds as follows. Section 2 derives the relationship between taxation and the long run allocation of labor in a search theoretic model. Section 3 describes the estimation of the model. Section 4 uses the estimated model to quantify the distortions of taxation on the allocation of labor and derives the mobility-based elasticity of taxable income. Sensitivity analyses are reported in Section 5, and Section 6 concludes.

2 Theory

In this section we set up a model of job search, labor allocation and taxation. It is an on-the-job search model similar to Burdett (1978) and Christensen et al. (2005), but with discrete wage distribution, zero discounting and including non-linear taxation. The assumption of a discrete wage distribution gives a direct correspondence between the theoretical model and the empirical implementation. The assumption of no discounting simplifies the welfare analysis of tax reforms considerably since we only need to consider the influence of a reform on the long run, steady state outcome to determine its welfare implications.

2.1 Fundamentals of the model

We consider a worker who is either employed in one out of a number of potential employment relationships/jobs ($i = 1, \dots, N$) paying different wages, or without any job ($i = 0$). When employed the worker may lose his job due to exogenous, random job destruction shocks. The worker may engage in search activity to find a better paid job when employed, and to find an employment opportunity when not employed. Thus, search leads to job mobility and labor reallocation.

In job i , the worker earns income w_i , where $w_1 < \dots < w_N$. In the non-employment state, $i = 0$, the worker obtains an income equivalent w_0 of leisure and home production. The net-of-tax income in job i equals $w_i - T_i$, where T_i is a tax-benefit function that includes all taxes and transfers, and $-T_0$ is the net benefit in the non-employment state. We define marginal tax rates as $m_i = (T_{i+1} - T_i) / (w_{i+1} - w_i)$ for $i = 1, \dots, N-1$, and $m_0 = (T_1 - T_0) / w_1$, where it is assumed

that $0 \leq m_i < 1$ for all i . We disregard low-wage job offers that would never be accepted by the worker and assume $w_1 - T_1 \geq w_0 - T_0$.

In state i , the worker decides on a search effort level $s_i \geq 0$. The cost of search effort is given by $c(s)$, where $c(0) = c'(0) = 0$, $c' > 0$, and $c'' > 0$. The worker receives job offers with Poisson frequency λs_i , where λ is a positive parameter. The wage rate of a job offer, w_j , is drawn from a given discrete probability distribution with probability p_h of w_h for $h = 1, \dots, N$. The cumulative distribution is $P_i = \sum_{j=1}^i p_j$. The worker may either accept the job offer and shift to job j , or reject the offer and stay in state i . Since the effort cost function $c(s)$ and the job offer distribution P are independent of the initial state i , it follows that a worker will accept a job offer if $w_j \geq w_i$. A worker who does not receive a job offer stays at state i .

When employed, the worker is exposed to a job destruction shock representing changing economic circumstances. At the Poisson frequency δ , the worker loses his current job, becomes unemployed and must wait for new job offers. The total job separation rate from job i is then $\delta + \lambda s_i (1 - P_i)$. If unemployed, the worker receives job offers drawn from the P -distribution with Poisson frequency λs_0 , and accepts any job offer w_j because of the assumption $w_j - T_j \geq w_0 - T_0$.

The income net of taxes and search cost in state i equals $w_i - T_i - c(s_i)$, and the worker decides on search effort in all states to maximize intertemporal utility defined as the expected present value of income net of taxes and search costs.

2.2 Optimal search behavior

A given search behavior, s_0, s_1, \dots, s_N , results in a certain steady state distribution over states $i = 0, 1, \dots, N$, determining the share of time the worker will be non-employed, $i = 0$, and in jobs $i = 1, \dots, N$, respectively, in the long run. With many identical workers this steady state distribution would also be the long run distribution of workers over states at any given time. Let u be the steady state rate of non-employment, and let g_i be the steady state fraction of employment in job $i = 1, \dots, N$, where $\sum_{j=1}^N g_j = 1$. The cumulative distribution is $G_i = \sum_{j=1}^i g_j$.

The assumption of no discounting implies that the intertemporal utility W of the worker equals the expectation of income net of tax and search cost with respect to the steady state distribution of employment and wages:

$$W = u [w_0 - T_0 - c(s_0)] + (1 - u) \sum_{j=1}^N [w_j - T_j - c(s_j)] g_j. \quad (1)$$

The worker maximizes W with respect to $\{s_0, \dots, s_N\}$. The first-order conditions are

$$\frac{dW}{ds_0} = -u \cdot c'(s_0) + \frac{\partial u}{\partial s_0} \left[w_0 - T_0 - c(s_0) - \sum_{j=1}^N [w_j - T_j - c(s_j)] g_j \right] = 0, \quad (2)$$

$$\frac{dW}{ds_i} = -(1-u) c'(s_i) g_i + (1-u) \sum_{j=1}^N [w_j - T_j - c(s_j)] \frac{\partial g_j}{\partial s_i} = 0 \text{ for } i = 1, \dots, N. \quad (3)$$

To find the partial derivatives $\partial u / \partial s_0$ and $\partial g_j / \partial s_i$, we must determine how the steady state distribution depends on behavior. The flow out of non-employment equals $u \lambda s_0$, and the flow into non-employment is $(1-u) \delta$. In steady state these flows are identical giving

$$u = \frac{\delta}{\lambda s_0 + \delta}. \quad (4)$$

Consider the set of jobs $I = \{1, \dots, i\}$. According to the steady state distribution, the worker will spend the fraction $(1-u) G_i$ of time in set I . The flow into I comes from unemployment and equals $u \lambda s_0 P_i$. The flow out of I to unemployment is $\delta (1-u) G_i$, and the flow to higher wages is $(1-u) \sum_{j=1}^i g_j \lambda s_j (1-P_i)$. The condition that flow out equals flow in is

$$\delta (1-u) G_i + (1-u) \sum_{j=1}^i \lambda s_j g_j (1-P_i) = u \lambda s_0 P_i,$$

which together with equation (4) gives

$$G_i = P_i - \frac{1}{\delta} (1-P_i) \sum_{j=1}^i \lambda s_j g_j, \quad (5)$$

showing that G stochastically dominates P because search on-the-job implies that the worker climbs up the income ladder. Note also that an increase in s_i implies that the post-change distribution stochastically dominates the pre-change distribution.

From the first-order conditions (2) and (3) and the relationships (4) and (5) characterizing the steady state distribution, we may derive the following formulas for the optimal search behavior (Appendix A):

$$s_N = 0, \quad (6)$$

and

$$c'(s_i) = \lambda \frac{(w_{i+1} - T_{i+1}) - (w_i - T_i) - [c(s_{i+1}) - c(s_i)] + s_{i+1} c'(s_{i+1}) - s_i c'(s_i)}{\delta} (1-P_i) + c'(s_{i+1}), \quad (7)$$

for $i = 0, 1, \dots, N-1$. This gives a recursive determination of optimal effort levels. Given $s_N = 0$, equation (7) implicitly yields a solution for s_{N-1} , and given this s_{N-1} , it gives a solution for s_{N-2} and so on.

2.3 Effects of tax reforms on search behavior, labor allocation and welfare

To illustrate how taxes work in this setting, we now consider small tax reforms formalized as infinitesimal changes in the tax liabilities, dT_i , for all $i = 0, \dots, N$. It is seen directly from equation (7) that search behavior only depends on taxation through $T_{i+1} - T_i$, and therefore only on the marginal tax rates since $T_{i+1} - T_i = m_i(w_{i+1} - w_i)$.

From $s_N = 0$ follows $ds_N = 0$. To obtain the effects on search effort levels for $i = 0, 1, \dots, N-1$, we differentiate equation (7). This gives

$$\left[\frac{\delta}{\lambda(1-P_i)} + s_i \right] ds_i c''(s_i) - \left[\frac{\delta}{\lambda(1-P_i)} + s_{i+1} \right] ds_{i+1} c''(s_{i+1}) = -dm_i(w_{i+1} - w_i). \quad (8)$$

Inspection of equation (8) shows that if the marginal tax rate is increased at income level w_j for some $j \leq N-1$, and unchanged everywhere else, then search effort at all wage levels up to and including j decreases, while effort is unchanged at higher income levels. Hence, irrespective of the income level at which the change occurs, the worker's expected (pre tax) income decreases. This is in contrast to a standard hours-of-work, labor supply model, where the worker's behavior and hence income is only affected if the worker happens to have a pre-change income level exactly where the marginal tax rate is changed.

In order to understand the empirical challenges of identifying the behavioral responses to taxation in a setting with search and labor mobility, consider a change of the tax system where first all marginal tax rates are identical, and then the marginal tax rates for income levels above a certain threshold are raised to a new, higher value, thus introducing tax progression with a kink-point in the tax schedule. In a standard hours-of-work, labor supply model, all workers with income below the kink-point will keep hours and income unchanged, while all workers with income above it will decrease hours to an extent that will keep income at or above the kink-point. According to such a model, the tax experiment would split workers into a control group of unaffected (low income) workers and a treatment group of affected (high income) workers, and the post change tax system would make workers pile up at the income level of the kink-point thereby creating a discontinuity in the density function. This implies that it is possible to identify empirically the behavioral responses by comparing treatment and control groups using difference-in-difference type techniques or by exploiting the bunching at the kink-point, as done in the literature surveyed in the Introduction.

In the model studied here, the experiment would affect behavior at all income levels, also below the kink-point. Hence, although the model does create an income distribution with some workers above and some workers below the kink-point, these groups would not constitute a treatment and

a control group, respectively.⁶ Moreover, there would not be any bunching at the kink-point of the new tax system because search intensities decrease everywhere, and not just above the kink point. This explains why these two commonly used empirical methods cannot be expected to identify the behavioral effects of taxation on labor mobility and, as a consequence, the tax distortion of the allocation of labor.

Next, we look at the impact of taxation on economic efficiency. The excess burden or dead-weight loss of taxation is defined as

$$D = \tilde{W} - W - R, \quad (9)$$

where W is the private welfare of equation (1) at the given tax system, \tilde{W} is private welfare in the absence of taxation, and R is the tax revenue defined as

$$R = u \cdot T_0 + (1 - u) \sum_{i=1}^N T_i g_i, \quad (10)$$

where u and g_1, \dots, g_N are the distributional parameters at the given tax system.

In Section 3 and Section 4, we quantify the welfare effects of taxation by estimating the structural parameters of the model and then compute directly the change in W , R and D caused by different types of tax reforms. To illustrate the main determinants of these welfare effects, we first analyze theoretically the impact of small reforms.

The change in the dead-weight loss of taxation from a small change in the tax system is equal to (see Appendix B)

$$dD = \left[\sum_{i=1}^N (T_i - T_0) g_i \right] \frac{\partial u}{\partial s_0} ds_0 - (1 - u) \sum_{j=1}^N \sum_{i=j}^N (T_i - T_j) \frac{\partial g_i}{\partial s_j} ds_j, \quad (11)$$

where ds_i is the change in search effort levels. This result reflects the general insight from the theory of taxation that the efficiency effects of small policy reforms are given simply by the behavioral revenue effects.⁷ To see that this is the case, consider a reform that reduces marginal taxes and increases search intensities at all states except N . This will move a probability mass $g_i |\partial u / \partial s_0| ds_0$ from unemployment to job i , which will increase tax revenue by $[T_i - T_0] g_i |\partial u / \partial s_0| ds_0$. This explains the first term on the right hand side of (11). In the same vein, the increased search effort in job $j \in \{1, \dots, N - 1\}$ will move probability mass $(1 - u) |\partial g_i / \partial s_j| ds_j$ from job j to job i where $i > j$, implying an indirect tax effect of $[T_i - T_j] (1 - u) |\partial g_i / \partial s_j| ds_j$. This explains the second

⁶In a homogeneous agent model there could never be a treatment and a control group, but in a heterogeneous agent version of the mobility model, where wage differences were partly due to differences in ability and partly due to differences in the workers' positions in the mobility cycle, the distinction between treatment and control group would still not be clear. For instance, workers with current income (just) below the kink-point could be in that position either due to relatively low ability or due to a temporarily low position in the mobility cycle. In the first case they would not be affected by the tax experiment, in the second case they would.

⁷This insight applies to any model where individuals optimize and the only source of inefficiency is distortionary taxation (Kleven and Kreiner, 2005).

term on the RHS of (11). All in all, the behavioral effect on tax revenue comes from the way the reform induces people to move from states with lower net taxes to states with higher net taxes.

If the initial tax system is linear, $T_i = T_0 + m \cdot w_i$ for $i \geq 1$, and we consider a uniform change dm in all marginal tax rates, then the formula (11) simplifies to (Appendix C)

$$\frac{dD}{Y} = \frac{m}{1-m} \cdot \varepsilon \cdot dm, \quad (12)$$

where

$$Y = (1-u) \sum_{j=1}^N w_j g_j \quad (13)$$

is taxable income (average earned income or the GDP) and

$$\varepsilon = \frac{\partial Y}{\partial(1-m)} \frac{1-m}{Y} \quad (14)$$

is the elasticity of taxable income with respect to the net-of-tax rate, $1-m$. Equation (12) is a standard elasticity formulation of the welfare effect where in our context ε is a ‘mobility elasticity’: The percentage increase in income caused by intensified search and thereby increased labor mobility for a one percent increase in the net of tax rate.

3 Empirical implementation

This section first describes the institutional features of the Danish labor market and our matched employer-employee data. Next, we outline some identification assumptions, and finally, the structural estimation approach is described. We follow the approach of Christensen et al. (2005) and estimate directly the wage distribution of job offers, P_i , the initial tax-benefit function, T_i , the job offer probability parameter, λ , the mobility effort cost function, $c(s)$, and the job destruction rate, δ .

3.1 The Danish labor market

Unlike many continental European countries, Denmark has a very flexible labor market (see, e.g., Botero et al. 2004). Employment protection is weak and turnover rates and average tenure are in line with those of the Anglo-Saxon countries. While most workers are in unions, wage formation has increasingly become more flexible as wage bargaining has been decentralized from the sector level to the firm level. In the late 1980s most wage contracts in the private labor market were negotiated at the sector level, but in our base sample year, 2005, less than 20% of the labor market was still covered by centralized wage bargaining. Instead wage contracts are negotiated either exclusively at the firm level or as a wage floor negotiated at the sector level combined with subsequent negotiations at the firm level. Dahl et al. (2012) shows that this decentralization

process has increased wage dispersion in the Danish labor market such that wages to a larger degree reflect local firm-level conditions.

3.2 Data description and empirical identification assumptions

We use data for the full population of workers and workplaces in Denmark for the years 2004-2006. The data are drawn from the Integrated Database for Labour Market Research (IDA), which is a longitudinal employer-employee register based data set maintained by Statistics Denmark. Workers and workplaces are linked and uniquely identified over time such that worker transitions in and out of the workplace may be tracked.⁸ In a specific week in each of the years 2004, 2005 and 2006, we know which workers were employed at each workplace.

In the following we focus on all 18-65 year old workers employed in private sector workplaces in 2005. For each workplace we register its size (the number of employees in 2005), the average wage rate of the employees, the number of “stayers” (the workers employed at the workplace in 2005 who are still employed at it in 2006), the number of “new hires” (the workers employed at the workplace in 2005 who were not employed there in 2004), and the number of “new hires from non-employment” (the workers employed at the workplace in 2005 who were not employed at any workplace in 2004, thus coming from unemployment or non-participation).

The wage rate of a worker is calculated as annual labor income (including mandatory pension fund payments) divided by annual hours. To reduce the impact of potential measurement errors in the calculation of workplace average wages, we exclude all observations in the top percentile (wage rates above DKK 714) in the wage distribution. In addition, wages below DKK 90 are dropped. A cut-off wage rate w_1 of DKK 90 corresponds approximately to the minimum wage floor negotiated at the sector level in 2005. This is also evident from the empirical wage distribution which exhibits a hump around DKK 90. In Section 5 on robustness, we report the results from a sensitivity analysis where the model is estimated with different cut-offs in the bottom of the wage distribution.

Following Christensen et al. (2005), we associate to each worker at a workplace the workplace’s average hourly wage rate. This average wage rate then corresponds to the wage w_i in job i of the theoretical model. The reason for this approach is that it is the firm component in wages that matters for job-to-job mobility in on-the-job search models. Under the assumption that worker and firm components in the wage are independent, firm-level average wages correspond to the firm component plus noise.

In the resulting data set, we have 126,400 workplaces with an average size of 11 employees, an

⁸In IDA a workplace is an establishment defined by a geographical location. Most often one firm has only one workplace, but few (larger) firms have several and thus split up in several workplaces in the data.

average number of stayers of 7.7, an average wage rate of DKK 197.6, an average number of new hires of 3.3, with 0.6 of these being new hires from non-employment, cf. Table 1.

TABLE 1. WORKPLACE SUMMARY STATISTICS

Variables	Mean	Median	Min	Max
Size	11.0	4.0	1.0	4141.0
Stayers	7.7	2.0	0.0	3677.0
Wage	197.6	183.2	90.0	714.1
New hires	3.3	1.0	0	2233.0
- from non-employment	0.6	0.0	0.0	158.0
# workplaces		126,400		

In the empirical implementation we need data for marginal tax rates. We use the empirical tax function for Denmark for 2006 constructed by Kleven and Kreiner (2006), which provides effective marginal tax rates as a function of annual labor income. The effective marginal tax rate at a given annual income level includes income taxes, indirect taxes and the marginal claw-back on any benefits that the individual is receiving. To apply this tax function to our 2005 data, we first deflate it by the overall growth rate in the average annual labor income between 2005 and 2006. Next, we convert the hourly wage rates in the data into annual labor incomes by multiplying it with the median annual hours worked in the data (1,660 hours). In this way, we assign an effective marginal tax rate to each hourly wage from DKK 90 up to DKK 714 as illustrated in Figure 1. These correspond to m_1, \dots, m_N . For m_0 we use in our baseline estimation the value 0.59, which is the marginal tax rate of the lowest possible wages around DKK 90. In the sensitivity analysis in Section 5, we consider also a higher value of m_0 in order to account for the loss of unemployment benefits when entering the labor market.

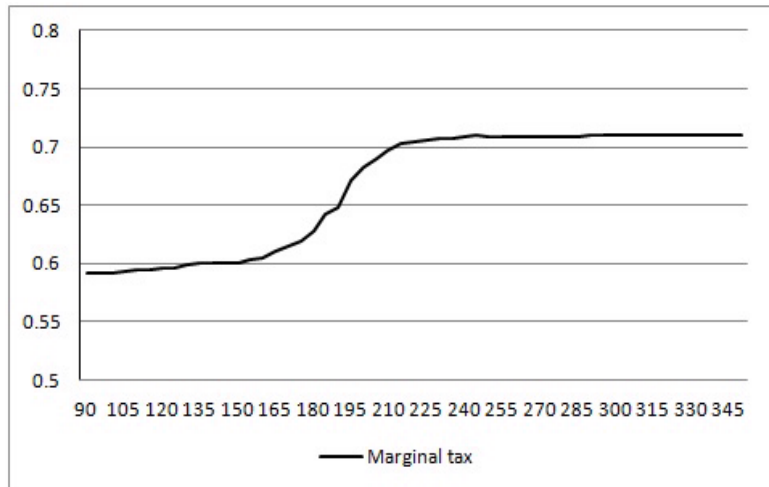


Figure 1: Effective marginal tax rates as a function of hourly wage rates.

The effective marginal tax rates are sufficient information to estimate the model and to perform subsequent policy analysis because (as mentioned) behavior only depends on taxation through the marginal tax rates. We can therefore construct a tax function from the marginal tax rates by choosing T_0 arbitrarily and let $T_1 = T_0 + m_0 w_1$ and $T_{i+1} = T_i + m_i(w_{i+1} - w_i)$, $i = 1, \dots, N - 1$.

We assume that workers are indifferent between working at the lowest wage level in the wage distribution and not working at all, $w_1 - T_1 = w_0 - T_0$, which implies $w_0 = w_1 - (T_1 - T_0) = (1 - m_0)w_1$. With $m_0 = 0.59$ and $w_1 = 90$, we obtain $w_0 = 36.9$. Hence, the value of leisure, home production etc. corresponds to an hourly wage of DKK 36.9.

Again following Christensen et al. (2005), the wage offer distribution, P_i , is defined as the distribution of firm-specific average wages for persons entering firms in 2005 from non-employment in 2004. Thus, P_i is constructed by weighting the firm-specific wage of firm i by the fraction of all workers hired from non-employment by firm i . This distribution may be compared to the distribution of earned average wages for all workers, and as expected from equation (5) the earned wage distribution is stochastically dominating the offered wage distribution, cf. Figure 2.

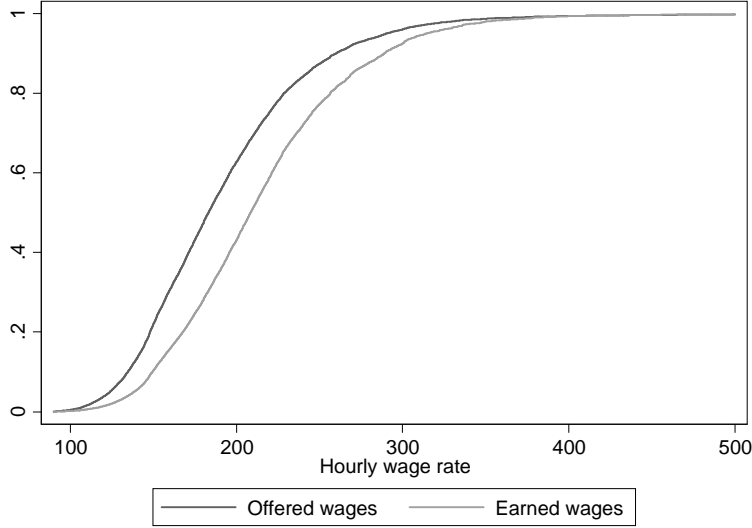


Figure 2: Offered wage and earned wage distributions

Finally, to identify the model, we assume that the mobility cost function is isoelastic:

$$c(s) = c_0 \frac{s^{1+1/\gamma}}{1 + 1/\gamma}, \quad (15)$$

with the parameters c_0 and γ .

3.3 Structural Estimation

Using the cost function (15), we may rewrite equation (7) determining the optimal search levels as

$$(\lambda s_i)^{1/\gamma} = \frac{\alpha [(w_{i+1} - T_{i+1}) - (w_i - T_i)] + \frac{1}{1+\gamma} [(\lambda s_{i+1})^{1+1/\gamma} - (\lambda s_i)^{1+1/\gamma}]}{\delta} (1 - P_i) + (\lambda s_{i+1})^{1/\gamma}, \quad (16)$$

where $\alpha = \frac{\lambda^{1+1/\gamma}}{c_0}$. The levels of search effort determine the job separation rates which, for a firm paying the wage level w_h , equals

$$d_h = \delta + \lambda s_h (1 - P_h), \quad (17)$$

where s_h is the optimal mobility effort at wage w_h .

The key parameters α , δ and γ are estimated in the following way. For given values of α , δ and γ (keeping fixed the set of w_i and P_i and T_i), and starting with $\lambda s_N = 0$, equation (16) determines recursively $\lambda s_{N-1}, \dots, \lambda s_1, \lambda s_0$. This means that equation (16) yields a functional relationship from the triplet (α, δ, γ) to the effort levels $(\lambda s_0, \lambda s_1, \dots, \lambda s_N)$, and then to the hazard rates d_h of equation (17). In the data, we observe for each workplace h the number of employees n_h in 2005, and the number of stayers x_h from 2005 to 2006. Given the separation rate d_h , the Poisson probability of zero separations within one period is e^{-d_h} , which is the probability of staying. Hence, the probability of x stayers within a year is given by the binomial distribution function

$$\Pr(x_h = x | n_h, d_h) = \binom{n_h}{x} e^{-d_h x} (1 - e^{-d_h})^{n_h - x},$$

and the probability of observing x_1, \dots, x_H over all firms is

$$L(\delta, \alpha, \gamma) = \prod_{h=1}^H \binom{n_h}{x_h} e^{-d_h x_h} (1 - e^{-d_h})^{n_h - x_h},$$

where $d_h = \delta + \lambda s_h (1 - P_h)$ depends directly on δ and, through λs_h , on (α, δ, γ) . The log likelihood equals

$$\ln L(\delta, \alpha, \gamma) = \sum_{h=1}^H \left[\ln \binom{n_h}{x_h} - d_h x_h + (n_h - x_h) \ln(1 - e^{-d_h}) \right].$$

The triplet (α, δ, γ) is estimated by maximization of $\ln L(\alpha, \delta, \gamma)$ taking into account the functional relationship from (α, δ, γ) to separation rates.

The resulting parameter estimates of our baseline estimation are displayed in Table 2, while Section 5 reports the outcomes of using alternative specifications.⁹ The parameter estimates are close to the findings of Christensen et al. (2005). The estimate $\gamma = 1.268$ implies a search cost function that is not far from quadratic.

⁹The Gauss program for the computation of the parameters are available from the authors upon request.

TABLE 2. PARAMETER ESTIMATES

	Point estimate	Standard error
α	0.0098	0.0001
δ	0.2781	0.0001
γ	1.2682	0.0071

The estimation produces a final vector of Poisson intensities $(\lambda_{s_0}, \lambda_{s_1}, \dots, \lambda_{s_N})$, and hence from equations (4) and (5), a rate of non-employment u , and a wage distribution G . The rate of non-employment equals 0.295, which to a large extent reflects the high job destruction rate, δ . The rate of non-employment is very high compared to realistic unemployment levels. However, the estimate is more reasonable if including all people of working age who are out of employment temporarily, including those who have left the work force temporarily for various reasons, e.g., parental leave, supplementary education etc. In the sensitivity analysis of Section 5, we consider a modification of the job destruction process that brings the estimated rate of non-employment close to levels compatible with a rate of structural unemployment. See also the discussion in Christensen et al. (2005).

To evaluate the models ability to reproduce the empirical wage distribution we can compare the computed g -distribution with the wage distribution in the data. As is evident from Figure 3 there is a close connection between the two.

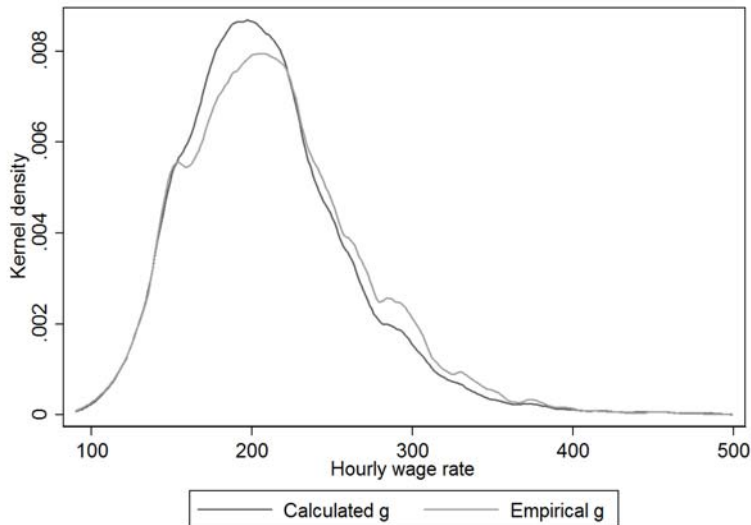


Figure 3: Calculated and empirical wage distribution

4 Quantifying the Tax Distortions of the Allocation of Labor

This section uses the estimated model to measure the tax distortions of the allocation of labor. This is done by analyzing the effects of hypothetical tax reforms that change the tax system from one system T with marginal tax rates m_i (for example the initial tax system) to another tax system \hat{T} , involving new marginal tax rates \hat{m}_i . The changes in the tax liabilities and marginal tax rates are $\Delta T_i = \hat{T}_i - T_i$, and $\Delta m_i = \hat{m}_i - m_i$, respectively, where $\Delta T_{i+1} - \Delta T_i = \Delta m_i (w_{i+1} - w_i)$ for all i .

Since our focus is on the effects of marginal taxation of the employed on the mobility and allocation of labor, we only consider changes of the marginal tax rates, m_1, \dots, m_{N-1} , along the intensive margin, while the participation tax along the extensive margin, $m_0 = (T_1 - T_0)/w_1$, is unchanged from the initial tax system in all of the experiments.¹⁰

Given the structural parameters (and P_i), we may compute the optimal Poisson search intensities λs_i and $\lambda \hat{s}_i$ associated with the tax system before and after the reform, respectively, using equation (16). From these search intensities, it is possible to derive pre-reform and post-reform rates of non-employment, u and \hat{u} , from equation (4), wage distributions, g and \hat{g} , from equation (5), aggregate income levels, Y and \hat{Y} , from equation (13), private welfare levels, W and \hat{W} , from equation (1), where the search costs $c(s_i)$ are computed by inserting $c_0 = \lambda^{1+1/\lambda}/\alpha$ into equation (15), and tax revenues, R and \hat{R} from equation (10). Given this, it is possible to compute the change in GDP, $\Delta Y = \hat{Y} - Y$, in aggregate welfare, $\Delta(W + R) = \Delta W + \Delta R$, and in the excess burden of taxation, $\Delta D = -\Delta(W + R)$.

The mechanical change in tax revenue measures the change in government revenue before behavioral changes and equals

$$\Delta M = (1 - u) \sum_{i=1}^N \Delta T_i g_i, \quad (18)$$

where u and g_i are associated with the pre-reform tax system. We can then define the behavioral change in tax revenue as

$$\Delta B = \Delta R - \Delta M. \quad (19)$$

The change in excess burden of the transition from T to \hat{T} is ΔD . When measured relative to the mechanical change in tax revenue, $\Delta D/\Delta M = -\Delta(W + R)/\Delta M$, we obtain, for a tax reduction, the welfare gain per unit of mechanic loss in tax revenue. We may also define the

¹⁰This is also an analytical necessity. The condition $w_1 - T_1 = w_0 - T_0$ must hold both before and after a reform experiment. Since w_0 is a deep parameter (inferred from observation of w_1 and $T_1 - T_0$), a change in m_0 would require a change in $w_1 = w_0/(1 - m_0)$. To compute the consequences would require a new domain of the P -distribution (from the new w_1 to w_N), but P and its domain are exogenous in the analysis and can (in principle) only be observed for the existing w_1 . In all the experiments we conduct, T_0 and T_1 and hence m_0 are kept unchanged from the initial tax system.

degree of self-financing, $-\Delta B/\Delta M$, which in the case of a tax reduction measures the fraction of the mechanical loss in tax revenue that returns from positive behavioral effects. As explained in Subsection 2.3, if the changes Δm_i in the tax system are small, then $\Delta D \approx -\Delta B$, implying that the marginal excess burden and the degree of self-financing become approximately identical, $\Delta D/\Delta M \approx -\Delta B/\Delta M$.

4.1 The marginal excess burden and the elasticity of taxable income

First, we consider a small reform of the initial tax system where the marginal tax rates are reduced to create a one percent increase in the net-of-tax rates, $1 - m_i$, everywhere along the intensive margin, that is, $\Delta m_i = -\Delta(1 - m_i) = -0.01(1 - m_i)$, and $\hat{m}_i = m_i + \Delta m_i$ for $i = 1, \dots, N - 1$. We then compute the elasticity of taxable income as $\varepsilon = (\Delta Y/Y) / [\Delta(1 - m_i)/(1 - m_i)] = 100 \cdot \Delta Y/Y$. In our context, this estimate of the elasticity of taxable income measures the percentage increase in the income of the employed due to intensified search and labor mobility for a one percent increase in the net-of-tax rate of all employed. The elasticity estimate we obtain is $\varepsilon = 0.30$.¹¹

Interestingly, we obtain a result very close to 0.30 if we use the formula (12) based on linear taxation to compute an approximate estimate of the elasticity. The welfare gain from the considered reform is estimated to be 0.18 percent of GDP, that is, $-\Delta D/Y = 0.0018$. The (g_i -weighted) average effective marginal tax rate for our sample is $2/3$, so on average the net-of-tax rate is 33.3 percent and therefore on average the absolute change in the marginal tax rate is 0.33 percentage points. From equation (12), we then get the approximate elasticity estimate: $\varepsilon = \frac{\Delta D}{Y} / \left(\frac{m}{1-m} \Delta m \right) = 0.0018 / (2 \cdot 0.0033) \approx 0.27$. The fact that this estimate, based on an assumption of linear taxation, is so close to the above estimate, that accounts for the non-linearity of the tax system, reflects that the Danish tax system is indeed close to linear, as is also seen from Figure 1.

We may also use the reform experiment to compute the marginal excess burden of taxation as described above. When measured in proportion to the mechanical loss of tax revenue, we obtain an estimate of 87 percent. A marginal reduction in the tax rates, financed by lump sum taxation, yields a welfare gain equal to 87 cent per dollar. Because we consider a small reform, this estimate also corresponds to the degree of self financing of the tax cut ($\Delta D/\Delta M \approx -\Delta B/\Delta M$). Thus, the increase in tax revenue from the behavioral response is 87 percent of the mechanical loss in tax revenue. The high marginal excess burden/degree of self financing reflects that effective marginal tax rates are high in Denmark (in 2006),¹² and that the estimated elasticity is reasonably high.

¹¹Smaller relative changes in the net-of-tax rate lead to almost the same elasticity estimate, showing that the considered reform is sufficiently small to approximate a marginal change in taxation.

¹²The Danish effective tax rates on labor income have been reduced in recent tax reforms, for example implying that the highest rate has dropped from 72 percent to 66 percent (Danish Economic Council, 2011).

4.2 The overall impact of marginal taxes on the allocation of labor and welfare

The structural approach enables us to consider tax systems that are not just marginally different from the existing tax system. In this subsection, we consider alternative tax systems $\hat{T}(m)$ indexed by m , where $\hat{T}_0(m)$ and $\hat{T}_1(m)$ are always the same as in the existing tax system, $\hat{T}_0(m) = T_0$ and $\hat{T}_1(m) = T_1$, and $\hat{T}_i(m) = T_1 + m(w_i - w_1)$ for $i = 2, \dots, N$. That is, we consider tax systems with the same participation tax as in the existing system and some constant marginal tax rate, m , that applies to all wage levels above w_1 . For m we consider different alternative values, 0, 0.01, 0.02, ..., 1. We analyze the effects of going from the existing tax system to specific alternative systems $\hat{T}(m)$, and the effects of small changes from each $\hat{T}(m)$ to $\hat{T}(m + 0.01)$, $m = 0, \dots, 0.99$.

First, we consider the effects of eliminating the progressivity of the existing tax system T along the intensive margin. We do this by considering the alternative tax system $\hat{T}(\bar{m})$, where \bar{m} is defined such that the mechanical change in tax revenue in equation (18) is zero when replacing the existing tax system T with the new system $\hat{T}(\bar{m})$. This gives $\bar{m} \approx 0.6275$. The effects of implementing this system in the estimated, structural model are described in the first row of Table 3. It shows that the elimination of progressivity has negligible influence on unemployment, increases GDP by 0.8 percent, and increases welfare by 0.5 percent of initial GDP. These small magnitudes reflect again that the existing tax system is not far from being linear.

Second, we study a reform that completely eliminates marginal taxation along the intensive margin by replacing the existing tax system T with $\hat{T}(0)$. The results in the third row of Table 3 show that the elimination of marginal taxation lowers non-employment by 13.3 percentage points and increases GDP by 28 percent. The welfare effect is 12.4 percent of initial GDP. Measured in proportion to the mechanical loss of tax revenue, the welfare gain is around one third, and the degree of self-financing is approximately 13 percent. A total welfare cost of taxation equal to 1/3 of the mechanical revenue change may seem small compared to the estimated marginal excess burden of almost 90 percent found in Subsection 4.1. However, as we show below this reflects the well-known fact that the marginal excess burden is increasing in the level of taxation.

Table 3 shows in the second row the effects of replacing the existing tax system T with $\hat{T}(0.3)$, which may roughly represent a transition from a Scandinavian system with high marginal taxes to a system along the lines of low-tax OECD countries such as the United States. This reduces the rate of non-employment by around 10 percentage points, increases GDP by almost 20 percent (relative to the Scandinavian GDP), and gives an aggregate efficiency gain measured in proportion to GDP of 9.9 percent. Thus, almost 80 percent of the efficiency loss from marginal taxation (9.7% divided by 12.4%) would be eliminated by shifting from a Scandinavian tax system to the system

of a low-tax OECD country according to these estimates (as always this should be balanced against the distributional impact of such a reform). The higher excess burden and degree of self-financing of this reform compared to the complete removal of marginal taxation reflects that both marginal tax distortions and the effects of behavioral responses on government revenue are higher at higher levels of taxation.

TABLE 3. EFFECTS OF IMPLEMENTING ALTERNATIVE NEW TAX SYSTEMS

Tax rate m	Unemployment Δu	GDP $\Delta Y/Y$	Welfare $-\Delta D/Y$	Excess burden $\Delta D/\Delta M$	Self-financing $-\Delta B/\Delta M$
\bar{m}	-0.0001	0.0076	0.005	-	-
0.3	0.093	0.193	0.099	0.526	0.390
0.0	-0.133	0.281	0.124	0.344	0.132

Understanding the relationship between the marginal excess burden, $\Delta D/\Delta M$, and the tax rate is key to understanding the relationship between some of our results. Figure 4 illustrates how the marginal excess burden depends on the initial tax rate in the estimated model. To construct the graph, we have analyzed small tax reforms from each $\hat{T}(m)$ to $\hat{T}(m + 0.01)$ for $m = 0, \dots, 0.99$. The graph shows an increasing and strictly convex relationship implying that the marginal excess burden of raising taxes is rather low at low levels of taxation and rather high at high levels of taxation. Note that the marginal excess burden is approximately equal to the degree of self-financing, $-\Delta B/\Delta M$, because we are considering small tax reforms, i.e., small changes in m . Thus, the graphs also shows that the Laffer rate (where $\Delta D/\Delta M \approx -\Delta B/\Delta M$ is equal to one) in the linear tax system is approximately 66 percent .

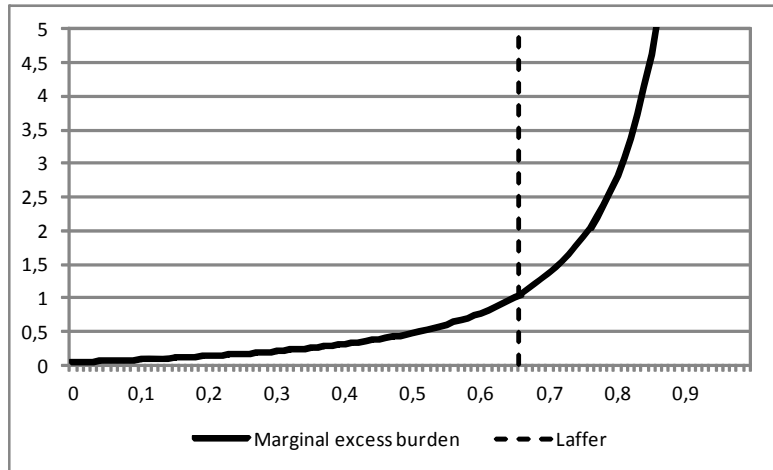


Figure 4: Marginal excess burden relative to mechanical revenue loss, $\Delta D/\Delta M$, (\approx degree of self financing, $-\Delta B/\Delta M$) for m going from 0 to 100 percent

5 Robustness

The standard errors of Table 2 are very small and we have not reported confidence intervals of our further estimations of, e.g., ε , but these are very narrow. Still our results could be sensitive to potential mis-specifications. In this section we analyze the robustness of our quantitative results for various alternative and relevant specifications.

First, as mentioned in Subsection 3.3, the estimated rate of non-employment is rather high, which reflects a high job destruction rate. To account for the possibility that some of the job destruction in the data is not “true” job destruction because of measurement error, we consider the case where there could be a “data job destruction rate” equal to some multiplum, $q\delta$, of the true job destruction rate, δ . The observed job separations of the estimation are governed by the “data separation rate”

$$d'_h = q\delta + \lambda s_h (1 - P_h), \quad (20)$$

but the optimal mobility efforts entering the estimation are governed by the true destruction rate δ , which is also involved in the determination of the rate of non-employment via (4). For the parameter q , we consider three values: $q = 1$, which is the baseline estimation, $q = 3$, which gives a true job destruction rate in line with other empirical evidence for Denmark by Rosholm and Svarer (2004),¹³ and $q = 6$, which gives a magnitude of the rate of non-employment around the level of the Danish structural unemployment rate, that may be considered as a lower estimate of the long run non-employment rate. For each q considered the full estimation is redone and then the tax experiment with a common one percent increase in the marginal net-of-tax rates along the intensive margin is redone to derive an elasticity of earned income. Table 4 reports our results; all standard errors are very small and not reported.

TABLE 4. ROBUSTNESS: ALTERNATIVE q

q	1	3	6
α	0.0098	0.0057	0.0041
δ	0.2781	0.0936	0.0470
γ	1.2682	1.8717	2.6141
u	0.295	0.126	0.068
ε	0.30	0.20	0.17
$\Delta D/\Delta M$	0.87	0.57	0.49

The fundamental parameter estimates are quite sensitive to q , but the derived elasticity is in all cases estimated to be in the interval [0.17; 0.30]. The welfare gain relative to the mechanical revenue loss reaches from 87 percent down to approximately 50 percent.

¹³Using different data than here, Rosholm and Svarer observe the employment status of a one percent representative sample of Danish workers week by week over almost 10 years. They estimate a Poisson job destruction rate on weekly basis of 0.0019. On annual basis this corresponds to a Poisson rate of approximately 0.0988 (= 52 times 0.0019), which is around one third of the rate estimated here (0.2781).

Second, while annual labor income is well measured,¹⁴ there may be a concern about the use of annual hours in the construction of our wage rate measure. Annual hours are measured with less precision as they are based on information from the mandatory pension fund ATP, which collects a relatively modest mandatory pension fund payment from all workers in the Danish labor market. The payment depends on the number of hours worked, and so hours may be imputed, but overtime work is not accounted for. This may be particularly problematic in our context where it is obviously of importance to have a precise measure of the change in wage rate associated with passing from one job to another. The measured changes could partly be due to changes in hours worked rather than only to payments per hour. To alleviate this concern we draw on the Wage Statistics Register (WSR or Lønstatistikken), which contains an alternative wage rate per hour reported by employers. This measure accounts explicitly for overtime work. The downside of it is a rather low coverage rate. The 126,400 workplaces in our sample employ in total 1.4 million workers with an observed IDA wage rate, but only 60% of these workers have a reported WSR wage rate.

Our strategy, intended to adjust the IDA wage rate for overtime work, is to estimate a relationship between the two wage variables for the workers with information about both measures. Given the IDA wage rate, we can then impute the overtime adjusted wage rate for the full sample based on this relationship. In order to obtain a reliable relationship we first trim the sample of IDA wages. We drop the top and bottom one percent of the observations, and we also eliminate IDA wage rates that are deemed of low quality by Statistics Denmark. This reduces the number of observations to 54% of the original sample of workers. The two variables are as expected highly correlated with a correlation coefficient of 0.87.

The next step is to estimate the relationship between the two measures. We use fractional polynomial regression, which is a very flexible way to estimate a relationship between two variables.¹⁵ This yields the following relationship

$$E(w_{WSR}|w_{IDA}) = 215.7 + 37.9 \left(\left(\frac{w_{IDA}}{100} \right)^2 - 5.1 \right) - 13.7 \left(\left(\frac{w_{IDA}}{100} \right)^2 \ln \left(\frac{w_{IDA}}{100} \right) - 4.1 \right).$$

Using this functional form, Figure 5 plots the predicted overtime adjusted wage rate against the IDA wage and also contains a 45°-line. The predicted wage rate is lower than than the IDA wage rate in most of the wage distribution, which is consistent with the existence of overtime work. At the bottom of the distribution the predicted wage rate is actually larger than the IDA wage rate indicating that here actual work hours may be shorter than official work hours.

¹⁴Annual labor income is based on tax return data. The Danish Tax Agency obtains third-party information from employers on nearly all labor income in Denmark, and tax evasion on labor income is almost zero (Kleven et al., 2011).

¹⁵We use the FRACPOLY routine in Stata.

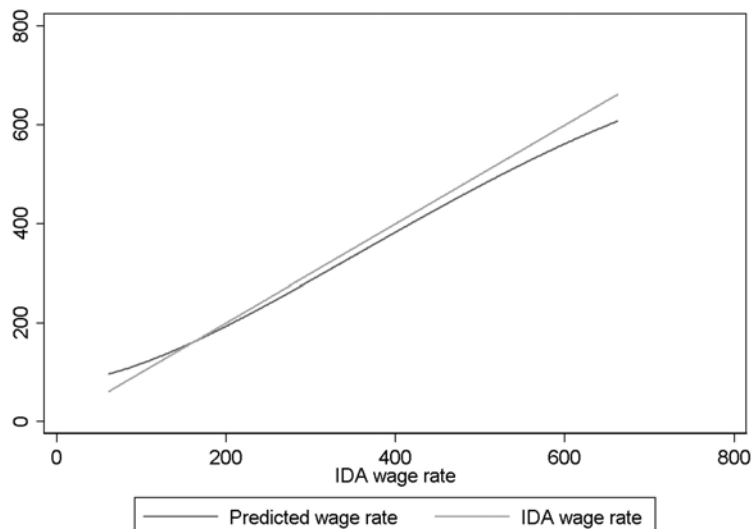


Figure 5. Predicted, overtime adjusted wage rate against IDA wage rate

The final step in this robustness exercise is to replace the IDA wage rate with the predicted overtime adjusted wage rate for all workers in the sample, calculate the average workplace wage rate, then estimate the model using this data set, and afterwards estimate once more the elasticity of taxable income. The full exercise is meant to eliminate hidden hours effects at job shifts and be left as closely as possible with a pure productivity effect on the hourly wage rate. Table 5, first column shows that neither the estimates of structural parameters, nor the implied elasticity change to any substantial degree with this alternative specification.

TABLE 5. ROBUSTNESS: VARIOUS ALTERNATIVE SPECIFICATIONS

	Overtime adj. wage rate	$w_1 = 80$	$w_1 = 100$	$m_0 = 0.82$
α	0.0120	0.0098	0.0095	0.0098
δ	0.2771	0.2770	0.2792	0.2781
γ	1.3519	1.2541	1.2412	1.2682
u	0.274	0.282	0.314	0.295
ε	0.29	0.28	0.31	0.30
$\Delta D/\Delta M$	0.84	0.76	0.99	0.97

Third, we consider two alternative values of the cutoff wage rate w_1 , DKK 80 and DKK 100. In each alternative case, w_0 is adjusted to preserve $w_0 = (1 - m_0)w_1$, and the P -distribution is reestimated on the relevant domain. Columns two and three of Table 5 reveal only small differences in the parameters and the estimates of the elasticity of taxable income ε lie in the narrow interval $[0.28; 0.31]$.

Fifth, the participation tax, $m_0 = 0.59$, was in the baseline estimation set at the level of the lowest marginal tax rates. The results of Kleven and Kreiner (2007) indicate that if the loss of

unemployment/social benefits is included then the participation tax rate on the increase in income from zero to the lowest possible, full time work income is 82 percent. It is not clear that this is the right participation tax rate since many of the non-employed will only be entitled to these benefits for a limited period. However, to explore the possible consequences, we may set m_0 equal to 0.82, keeping all other tax rates unchanged, and again adjust w_0 so that $w_0 = (1 - m_0)w_1$ etc. This means that the difference in net income between any two positions i and j is unchanged from the baseline estimation and hence the estimates of the structural parameters and the rate of non-employment as well as the elasticity of taxable income are also unchanged. The welfare effect is affected, however, since now a decrease of marginal tax rates will cause a larger behavioral increase in tax revenue since mobility effort is increased everywhere and the difference between T_0 and T_1 has increased, so there is a larger behavioral tax gain as non-employment is reduced. Column 4 of Table 5 shows that the estimate of the marginal excess burden (the degree of self financing) goes up to 97 percent, so in this case the considered tax experiment is close to a “Laffer change”.

6 Concluding remarks

In this paper, we apply a microeconomic, structural analysis to measure the distortionary effects of taxation on labor mobility and the long run allocation of labor. The analysis points to non-trivial effects of taxes on labor mobility. We estimate a mobility-based elasticity of taxable income with respect to the net-of-tax rate in the range 0.15-0.3. This is of the same magnitude as the estimates found in the empirical public finance literature using difference-in-difference type methods (Saez et al., 2012) and bunching at kink-points in the tax schedule (Chetty et al., 2011a) to measure the elasticity of taxable income. The aim of these methods is to capture all relevant behavioral responses to taxation, thereby obtaining a sufficient statistic for evaluating the efficiency effects of taxation. However, the methods fail to identify the effects of taxation on labor mobility. Our structural approach relies on a number of strong assumptions (independence of firm and worker components in wages, exogenous distribution of job opportunities, specific functional forms, no general equilibrium effects etc.) and exploits only cross-sectional variation across individuals for identification. Despite these weaknesses, the results indicate, in our view, that the empirical public finance literature may be overlooking sizable behavioral mobility effects when measuring the elasticity of taxable income.

A Derivation of equations (6) and (7)

The derivation proceeds in three steps. In Step 1 we use equation (5) to derive expressions for g_1, \dots, g_N , which are useful for computing the derivatives $\partial g_j / \partial g_i$ appearing in the first order conditions (3). In Step 2 we compute $\partial u / \partial s_0$ from (4) and $\partial g_j / \partial g_i$ from the expressions for g_j obtained in Step 1. Finally, in Step 3, we insert the expressions for $\partial u / \partial s_0$ and $\partial g_j / \partial g_i$ of Step 2 into the first order conditions (2) and (3), and arrive at (6) and (7).

A.1 Step 1

We first show that equation (5) implies

$$g_1 = p_1 \frac{\delta}{\delta + \lambda(1 - P_1) s_1}, \text{ and} \quad (\text{A-1})$$

$$g_i = \frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_i \prod_{j=1}^{i-1} \frac{\delta + \lambda(1 - P_{j-1}) s_j}{\delta + \lambda(1 - P_{j+1}) s_{j+1}}, \quad (\text{A-2})$$

for $i = 2, \dots, N$. From equation (5), we have

$$G_1 = P_1 - \frac{\lambda}{\delta} (1 - P_1) s_1 g_1.$$

This may be written as

$$g_1 = p_1 - \frac{\lambda}{\delta} (1 - P_1) s_1 g_1 = p_1 \frac{\delta}{\delta + \lambda(1 - P_1) s_1}$$

which is equation (A-1). From equation (5), we get

$$\delta G_i + \lambda(1 - P_i) \sum_{j=1}^i s_j g_j = \delta P_i,$$

and for $i \geq 2$, we have

$$\delta G_{i-1} + \lambda(1 - P_{i-1}) \sum_{j=1}^{i-1} s_j g_j = \delta P_{i-1}.$$

By subtracting the second from the first of the two expressions above and splitting up summations, we obtain for $i \geq 2$:

$$\delta g_i + \lambda \left\{ \left[\sum_{j=i+1}^N p_i \right] \left[\sum_{j=1}^{i-1} s_j g_j + s_i g_i \right] - \left[p_i + \sum_{j=i+1}^N p_i \right] \left[\sum_{j=1}^{i-1} s_j g_j \right] \right\} = \delta p_i,$$

or

$$\delta g_i + \lambda \left\{ \left[\sum_{j=i+1}^N p_i \right] s_i g_i - p_i \left[\sum_{j=1}^{i-1} s_j g_j \right] \right\} = \delta p_i,$$

or

$$[\delta + \lambda(1 - P_i) s_i] g_i = p_i \left(\delta + \lambda \sum_{j=1}^{i-1} s_j g_j \right)$$

or

$$g_i = p_i \frac{\delta + \lambda \sum_{j=1}^{i-1} s_j g_j}{\delta + \lambda(1 - P_i) s_i}. \quad (\text{A-3})$$

This is a recursive solution for g_i , $i = 2, \dots, N$ depending on s_1, \dots, s_i and g_1, \dots, g_{i-1} . We derive a solution for g_{i+1} only as a function of g_i and s_i and s_{i+1} . By leading (A-3), we get

$$g_{i+1} = p_{i+1} \frac{\delta + \lambda \sum_{j=1}^i s_j g_j}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} = p_{i+1} \frac{\delta + \lambda \sum_{j=1}^{i-1} s_j g_j + \lambda s_i g_i}{\delta + \lambda(1 - P_{i+1}) s_{i+1}}.$$

From (A-3), we have $\lambda \sum_{j=1}^{i-1} s_j g_j = g_i \frac{(\delta + \lambda(1 - P_i) s_i)}{p_i} - \delta$ which inserted in g_{i+1} gives

$$(\delta + \lambda(1 - P_{i+1}) s_{i+1}) g_{i+1} = p_{i+1} \left(\frac{\delta + \lambda(1 - P_i) s_i}{p_i} + \lambda s_i \right) g_i$$

which gives

$$g_{i+1} = \frac{p_{i+1}}{p_i} \frac{\delta + \lambda(1 - P_i) s_i}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} g_i. \quad (\text{A-4})$$

By expanding this relationship, we obtain

$$g_2 = \frac{p_2}{p_1} \frac{\delta + \lambda(1 - P_0) s_1}{\delta + \lambda(1 - P_2) s_2} g_1 = p_2 \frac{\delta + \lambda(1 - P_0) s_1}{\delta + \lambda(1 - P_2) s_2} \frac{\delta}{\delta + \lambda(1 - P_1) s_1},$$

where we have used (A-1). By expanding further, we get

$$\begin{aligned} g_3 &= \frac{p_3}{p_2} \frac{\delta + \lambda(1 - P_1) s_2}{\delta + \lambda(1 - P_3) s_3} g_2 \\ &= \frac{p_3}{p_2} \frac{\delta + \lambda(1 - P_1) s_2}{\delta + \lambda(1 - P_3) s_3} \frac{\delta + \lambda(1 - P_0) s_1}{\delta + \lambda(1 - P_2) s_2} \frac{\delta}{\delta + \lambda(1 - P_1) s_1} \\ &\dots \\ g_i &= \frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_i \prod_{j=1}^{i-1} \frac{\delta + \lambda(1 - P_{j-1}) s_j}{\delta + \lambda(1 - P_{j+1}) s_{j+1}}, \end{aligned}$$

which is (A-2).

A.2 Step 2

Here, we derive $\partial u / \partial s_0$ and $\partial g_j / \partial s_i$. Differentiation of equation (4) gives

$$\frac{\partial u}{\partial s_0} = -\frac{\delta \lambda}{(\lambda s_0 + \delta)^2} = -\frac{\lambda}{\lambda s_0 + \delta} u. \quad (\text{A-5})$$

We show below that differentiation of equations (A-1) and (A-2) give

$$\frac{\partial g_j}{\partial s_i} = \begin{cases} 0 & \text{for } j < i \\ -g_i \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} < 0 & \text{for } j = i \\ \frac{\lambda \delta p_i}{(\delta + \lambda(1 - P_{i-1}) s_i)(\delta + \lambda(1 - P_i) s_i)} g_j > 0 & \text{for } j > i \end{cases}. \quad (\text{A-6})$$

We start by changing indices in (A-2). This gives

$$g_j = \frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_j \prod_{h=1}^{j-1} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}}. \quad (\text{A-7})$$

We see directly that $\partial g_j / \partial s_i = 0$ for $j < i$.

By differentiating (A-7) with respect to s_i for $j = i$, we get

$$\begin{aligned}
\frac{\partial g_i}{\partial s_i} &= -\frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_i \left(\prod_{h=1}^{i-2} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \right) \frac{\delta + \lambda(1 - P_{i-2}) s_{i-1}}{(\delta + \lambda(1 - P_i) s_i)^2} \lambda(1 - P_i) \\
&= -\frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_i \left(\prod_{h=1}^{i-2} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \right) \frac{\delta + \lambda(1 - P_{i-2}) s_{i-1}}{\delta + \lambda(1 - P_i) s_i} \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} \\
&= -\frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_i \left(\prod_{h=1}^{i-1} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \right) \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} \\
&= -g_i \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i}
\end{aligned}$$

We now differentiate (A-7) with respect to s_i for $j > i$. For $i = 1$, we have

$$\begin{aligned}
\frac{\partial g_j}{\partial s_1} &= -\delta \frac{\lambda(1 - P_1)}{(\delta + \lambda(1 - P_1) s_1)^2} p_j \prod_{h=1}^{j-1} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \\
&\quad + \frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_j \frac{\lambda}{\delta + \lambda s_1} \prod_{h=1}^{j-1} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \\
&= \left[\frac{\lambda}{\delta + \lambda s_1} - \frac{\lambda(1 - P_1)}{\delta + \lambda(1 - P_1) s_1} \right] g_j
\end{aligned}$$

For $i > 1$, we get

$$\begin{aligned}
\frac{\partial g_j}{\partial s_i} &= \frac{\delta}{\delta + \lambda(1 - P_1) s_1} p_j \left(\frac{\lambda(1 - P_{i-1})}{\delta + \lambda(1 - P_{i-1}) s_i} - \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} \right) \prod_{h=1}^{j-1} \frac{\delta + \lambda(1 - P_{h-1}) s_h}{\delta + \lambda(1 - P_{h+1}) s_{h+1}} \\
&= \left(\frac{\lambda(1 - P_{i-1})}{\delta + \lambda(1 - P_{i-1}) s_i} - \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} \right) g_j \\
&= \frac{\lambda \delta (P_i - P_{i-1})}{(\delta + \lambda(1 - P_{i-1}) s_i) (\delta + \lambda(1 - P_i) s_i)} g_j \\
&= \frac{\lambda \delta p_i}{(\delta + \lambda(1 - P_{i-1}) s_i) (\delta + \lambda(1 - P_i) s_i)} g_j
\end{aligned}$$

This formula also holds for $i = 1$, which is easily seen by comparing the second equality to the previous expression for $\partial g_j / \partial s_1$.

A.3 Step 3

Consider optimal search effort at $i = N$. In this case, the condition (3) becomes

$$-c'(s_N) g_N + [w_N - T_N - c(s_N)] \frac{\partial g_N}{\partial s_N} = 0.$$

From equation (A-6), we have $\partial g_N / \partial s_N = 0$, because $P_N = 1$. Hence, $c'(s_N) = 0$ giving $s_N = 0$ in (6).

Next, we derive (7). We first consider $i \geq 1$. Note that in equation (3) one can let the summation start in i since from equation (A-6), $\partial g_j / \partial s_i = 0$ for $j < i$. By isolating $c'(s_i)$ from (3), we then obtain

$$c'(s_i) = \frac{[w_i - T_i - c(s_i)] \frac{\partial g_i}{\partial s_i} + \sum_{j=i+1}^N [w_j - T_j - c(s_j)] \frac{\partial g_j}{\partial s_i}}{g_i}.$$

By inserting the derivatives $\partial g_j / \partial s_i$ from (A-6), we get

$$\begin{aligned} c'(s_i) &= -[w_i - T_i - c(s_i)] \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} + \frac{\lambda \delta p_i}{(\delta + \lambda(1 - P_{i-1}) s_i) (\delta + \lambda(1 - P_i) s_i)} \times \\ &\quad \frac{1}{g_i} \sum_{j=i+1}^N [w_j - T_j - c(s_j)] g_j \end{aligned} \quad (\text{A-8})$$

or

$$\begin{aligned} c'(s_i) &= -[w_i - T_i - c(s_i)] \frac{\lambda(1 - P_i)}{\delta + \lambda(1 - P_i) s_i} + \frac{\lambda \delta p_i}{(\delta + \lambda(1 - P_{i-1}) s_i) (\delta + \lambda(1 - P_i) s_i)} \times \\ &\quad \frac{1}{g_i} \left([w_{i+1} - T_{i+1} - c(s_{i+1})] g_{i+1} + \sum_{j=i+2}^N [w_j - T_j - c(s_j)] g_j \right) \end{aligned} \quad (\text{A-9})$$

By leading equation (A-8), we obtain

$$\begin{aligned} c'(s_{i+1}) &= -[w_{i+1} - T_{i+1} - c(s_{i+1})] \frac{\lambda(1 - P_{i+1})}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} \\ &\quad + \frac{\lambda \delta p_{i+1}}{(\delta + \lambda(1 - P_i) s_{i+1}) (\delta + \lambda(1 - P_{i+1}) s_{i+1})} \frac{1}{g_{i+1}} \sum_{j=i+2}^N [w_j - T_j - c(s_j)] g_j \end{aligned}$$

or

$$\begin{aligned} \sum_{j=i+2}^N [w_j - T_j - c(s_j)] g_j &= \frac{(\delta + \lambda(1 - P_i) s_{i+1}) (\delta + \lambda(1 - P_{i+1}) s_{i+1}) g_{i+1}}{\lambda \delta p_{i+1}} c'(s_{i+1}) \\ &\quad + [w_{i+1} - T_{i+1} - c(s_{i+1})] \frac{\lambda(1 - P_{i+1}) (\delta + \lambda(1 - P_i) s_{i+1}) g_{i+1}}{\lambda \delta p_{i+1}} \end{aligned} \quad (\text{A-10})$$

By substituting (A-10) into (A-9) and multiplying by $[\delta + \lambda(1 - P_i) s_i] / \lambda$, we get

$$\begin{aligned} \frac{\delta + \lambda(1 - P_i) s_i}{\lambda} c'(s_i) &= -[w_i - T_i - c(s_i)] (1 - P_i) \\ &\quad + \frac{\delta p_i}{\delta + \lambda(1 - P_{i-1}) s_i} \frac{g_{i+1}}{g_i} [w_{i+1} - T_{i+1} - c(s_{i+1})] \\ &\quad + \frac{\delta p_i}{\delta + \lambda(1 - P_{i-1}) s_i} \frac{g_{i+1}}{g_i} \frac{(\delta + \lambda(1 - P_i) s_{i+1}) (\delta + \lambda(1 - P_{i+1}) s_{i+1})}{\lambda \delta p_{i+1}} c'(s_{i+1}) \\ &\quad + \frac{\delta p_i}{\delta + \lambda(1 - P_{i-1}) s_i} \frac{g_{i+1}}{g_i} [w_{i+1} - T_{i+1} - c(s_{i+1})] \frac{\lambda(1 - P_{i+1}) (\delta + \lambda(1 - P_i) s_{i+1})}{\lambda \delta p_{i+1}} \end{aligned}$$

From equation (A-4), we have

$$\frac{g_{i+1}}{g_i} = \frac{p_{i+1}}{p_i} \frac{\delta + \lambda(1 - P_{i-1}) s_i}{\delta + \lambda(1 - P_{i+1}) s_{i+1}}$$

By substituting this relationship into the above expression, we get

$$\begin{aligned}
& \frac{\delta + \lambda(1 - P_i) s_i}{\lambda} c'(s_i) = -[w_i - T_i - c(s_i)](1 - P_i) \\
& + \delta p_{i+1} \frac{1}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} [w_{i+1} - T_{i+1} - c(s_{i+1})] \\
& + \frac{1}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} \frac{(\delta + \lambda(1 - P_i) s_{i+1})(\delta + \lambda(1 - P_{i+1}) s_{i+1})}{\lambda} c'(s_{i+1}) \\
& + \frac{1}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} [w_{i+1} - T_{i+1} - c(s_{i+1})] \frac{\lambda(1 - P_{i+1})(\delta + \lambda(1 - P_i) s_{i+1})}{\lambda}
\end{aligned}$$

or

$$\begin{aligned}
& \frac{\delta + \lambda(1 - P_i) s_i}{\lambda} c'(s_i) = -[w_i - T_i - c(s_i)](1 - P_i) \\
& + \delta p_{i+1} \frac{1}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} [w_{i+1} - T_{i+1} - c(s_{i+1})] \\
& + \frac{(\delta + \lambda(1 - P_i) s_{i+1})}{\lambda} c'(s_{i+1}) \\
& + \frac{1}{\delta + \lambda(1 - P_{i+1}) s_{i+1}} [w_{i+1} - T_{i+1} - c(s_{i+1})] (1 - P_{i+1})(\delta + \lambda(1 - P_i) s_{i+1})
\end{aligned}$$

or

$$\frac{\delta + \lambda(1 - P_i) s_i}{\lambda(1 - P_i)} c'(s_i) - \frac{\delta + \lambda(1 - P_i) s_{i+1}}{\lambda(1 - P_i)} c'(s_{i+1}) = [w_{i+1} - T_{i+1} - c(s_{i+1})] - [w_i - T_i - c(s_i)]$$

or

$$\begin{aligned}
& \delta c'(s_i) + \lambda(1 - P_i) s_i c'(s_i) - (\delta + \lambda(1 - P_i) s_{i+1}) c'(s_{i+1}) \\
& = \lambda \{ [w_{i+1} - T_{i+1} - c(s_{i+1})] - [w_i - T_i - c(s_i)] \} (1 - P_i)
\end{aligned}$$

By adding $\lambda(1 - P_i) s_{i+1} c'(s_i)$ and subtracting $\lambda(1 - P_i) s_i c'(s_i)$ on both sides, we get

$$\begin{aligned}
& \delta c'(s_i) + \lambda(1 - P_i) s_{i+1} c'(s_i) - (\delta + \lambda(1 - P_i) s_{i+1}) c'(s_{i+1}) \\
& = \lambda \{ [w_{i+1} - T_{i+1} - c(s_{i+1})] - [w_i - T_i - c(s_i)] \} (1 - P_i) \\
& - \lambda(1 - P_i) s_i c'(s_i) + \lambda(1 - P_i) s_{i+1} c'(s_i)
\end{aligned}$$

or

$$\begin{aligned}
c'(s_i) & = \lambda \frac{[w_{i+1} - T_{i+1}] - [w_i - T_i] - [c(s_{i+1}) - c(s_i)] + (s_{i+1} - s_i) c'(s_i)}{\delta + \lambda s_{i+1} (1 - P_i)} (1 - P_i) \\
& + c'(s_{i+1})
\end{aligned} \tag{A-11}$$

Next, we show that (A-11) also applies for $i = 0$. From equations (2) and (A-5), we get

$$-c'(s_0) - \frac{\lambda}{\lambda s_0 + \delta} \left[w_0 - T_0 - c(s_0) - [w_1 - T_1 - c(s_1)] g_1 - \sum_{j=2}^N [w_j - T_j - c(s_j)] g_j \right] = 0$$

or

$$0 = -(\lambda s_0 + \delta) c'(s_0) - \lambda [w_0 - T_0 - c(s_0)] + \lambda [w_1 - T_1 - c(s_1)] g_1 \\ + \lambda \sum_{j=2}^N [w_j - T_j - c(s_j)] g_j$$

Equation (A-10) holds also for $i = 0$, because it has been led. With $P_0 = 0$ this gives

$$\sum_{j=2}^N [w_j - T_j - c(s_j)] g_j = \frac{(\delta + \lambda s_1)(\delta + \lambda(1 - P_1)s_1)g_1}{\lambda \delta p_1} c'(s_1) \\ + [w_1 - T_1 - c(s_1)] \frac{\lambda(1 - P_1)(\delta + \lambda s_1)g_1}{\lambda \delta p_1}$$

By inserting this sum in the above equation, we arrive at

$$0 = -(\lambda s_0 + \delta) c'(s_0) - \lambda [w_0 - T_0 - c(s_0)] \\ + \lambda [w_1 - T_1 - c(s_1)] \frac{\delta + (1 - P_1)\lambda s_1}{\delta p_1} g_1 \\ + \frac{(\delta + \lambda s_1)(\delta + \lambda(1 - P_1)s_1)g_1}{\delta p_1} c'(s_1)$$

Inserting g_1 from equation (A-1) gives

$$0 = -(\lambda s_0 + \delta) c'(s_0) - \lambda [w_0 - T_0 - c(s_0)] + \lambda [w_1 - T_1 - c(s_1)] \\ + (\delta + \lambda s_1) c'(s_1)$$

or

$$c'(s_0) = \lambda \frac{[w_1 - T_1 - c(s_1)] - [w_0 - T_0 - c(s_0)] + (s_1 - s_0) c'(s_0)}{\delta + \lambda s_1} \\ + c'(s_1)$$

This shows that (A-11) also holds for $i = 0$ with $P_0 = 0$.

From (A-11) follows by multiplying on both sides by $\delta + \lambda s_{i+1} (1 - P_i)$ etc.

$$c'(s_i) = \lambda \frac{[w_{i+1} - T_{i+1}] - [w_i - T_i] - [c(s_{i+1}) - c(s_i)] + s_{i+1} c'(s_{i+1}) - s_i c'(s_i)}{\delta} (1 - P_i) \\ + c'(s_{i+1}),$$

which is (7).

B Derivation of Equation (11)

We first insert W and R from equations (1) and (10) into the expression (9) for D . We then total differentiate this expression thereby obtaining the change in the dead-weight loss of taxation dD as a functions of the changes in tax liabilities dT_i , the induced changes in search effort ds_i , and

the changes in the distributional parameters du and dg_i . By arranging the terms appropriately, dD can be split into three components in a way well-known from the theory of taxation

$$dD = dD|_{\text{Mechanical}} + dD|_{\text{Behavior on utility}} + dD|_{\text{Behavior on tax revenue}},$$

where

$$dD|_{\text{Mechanical}} = u \cdot dT_0 + (1-u) \sum_{i=1}^N dT_i g_i - u \cdot dT_0 - (1-u) \sum_{i=1}^N dT_i g_i$$

is the mechanical effect on $R + W$, and

$$dD|_{\text{Behavior on utility}} = \sum_{i=0}^N \frac{\partial W}{\partial s_i} ds_i$$

is the effect on utility (W) from changed behavior and

$$dD|_{\text{Behavior on tax revenue}} = - \left(T_0 - \sum_{i=1}^N T_i g_i \right) du - (1-u) \sum_{i=1}^N T_i \cdot dg_i$$

is the effect on tax revenue from behavioral responses. The mechanical effect is obviously zero and, from the Envelope Theorem ($\partial W / \partial s_i = 0$), the behavioral effect on utility is also zero. Thus,

$$dD = dD|_{\text{Behavior on tax revenue}} = \left(\sum_{i=1}^N [T_i - T_0] g_i \right) \frac{\partial u}{\partial s_0} ds_0 - (1-u) \sum_{j=1}^N \sum_{i=j}^N T_i \frac{\partial g_i}{\partial s_j} ds_j,$$

where we have used that $\partial g_i / \partial s_j = 0$ for $i < j$ according to equation (A-6). Since g is a probability distribution, $\sum_{i=1}^N \frac{\partial g_i}{\partial s_j} ds_j = \sum_{i=j}^N \frac{\partial g_i}{\partial s_j} ds_j = 0$, using again $\partial g_i / \partial s_j = 0$ for $i < j$. By subtracting $T_j \sum_{i=j}^N \frac{\partial g_i}{\partial s_j} ds_j = 0$ on the RHS of the above equation, we get (11).

C Derivation of Equation (12)

For $T_i = T_0 + m \cdot w_i$, equation (11) becomes

$$dD = m \left(\sum_{i=1}^N w_i g_i \right) \frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial (1-m)} d(1-m) - (1-u) m \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial (1-m)} d(1-m)$$

or

$$dD = -\frac{m}{1-m} \left[\left(\sum_{i=1}^N w_i g_i \right) \frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial (1-m)} - (1-u) \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial (1-m)} \right] \frac{1-m}{Y} Y \cdot dm.$$

To establish (12), we need to show that

$$\varepsilon = - \left[\frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial (1-m)} \left(\sum_{i=1}^N w_i g_i \right) - (1-u) \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial (1-m)} \right] \frac{1-m}{Y} \quad (\text{C-12})$$

From the definition of Y , we compute

$$\frac{\partial Y}{\partial(1-m)} = - \left[\frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial(1-m)} \sum_{i=1}^N w_i g_i - (1-u) \sum_{j=1}^N \sum_{i=j}^N w_i \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \right] \quad (\text{C-13})$$

Consider the last term and split it up

$$\sum_{j=1}^N \sum_{i=j}^N w_i \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} = \sum_{j=1}^N \sum_{i=j+1}^N w_i \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} + \sum_{j=1}^N w_j \frac{\partial g_j}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \quad (\text{C-14})$$

Since from (A-6) a change in s_j only affects g_i for $i \geq j$, and since changes ∂g_i in g_i must add up to zero over all i , we have

$$\frac{\partial g_j}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} + \sum_{i=j+1}^N \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} = 0$$

implying that

$$\frac{\partial g_j}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} = - \sum_{i=j+1}^N \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)}.$$

Inserting into (C-14) gives

$$\begin{aligned} \sum_{j=1}^N \sum_{i=j}^N w_i \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} &= \sum_{j=1}^N \sum_{i=j+1}^N w_i \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} - \sum_{j=1}^N \sum_{i=j+1}^N w_j \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \\ &= \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \end{aligned}$$

Inserting this for the last term in (C-13) gives

$$\frac{dY}{\partial(1-m)} = - \left[\frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial(1-m)} \sum_{i=1}^N w_i g_i - (1-u) \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \right]$$

Hence,

$$\begin{aligned} \varepsilon &= \frac{dY}{\partial(1-m)} \frac{1-m}{Y} \\ &= - \left[\frac{\partial u}{\partial s_0} \frac{\partial s_0}{\partial(1-m)} \sum_{i=1}^N w_i g_i - (1-u) \sum_{j=1}^N \sum_{i=j+1}^N (w_i - w_j) \frac{\partial g_i}{\partial s_j} \frac{\partial s_j}{\partial(1-m)} \right] \frac{1-m}{Y}, \end{aligned}$$

which is indeed (C-12).

References

- [1] Botero, J.C., S. Djankov, R. La Porta, F. Lopez-De-Silanes and A. Shleifer (2004). "The Regulation of Labor", *Quarterly Journal of Economics* 119(4), 1339-1382.

- [2] Burdett, K. (1978). "A Theory of Employee Job Search and Quit Rates", *American Economic Review* 68, 212-220.
- [3] Chetty, R. (2009). "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance." *American Economic Journal: Economic Policy* 1, 31–52.
- [4] Chetty, R. (2012). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80(3): 969-1018.
- [5] Chetty, R., J. Friedman, T. Olsen and L. Pistaferri (2011a). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126, 749-804.
- [6] Chetty, R., A. Guren, D. Manoli and A. Weber (2011b). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review Papers and Proceedings* 101, 471-75.
- [7] Christensen, B. J., R. Lentz, D. T. Mortensen, G. R. Neumann and A. Werwatz, 2005. "On-the-Job Search and the Wage Distribution." *Journal of Labor Economics* 23, 31-58.
- [8] Dahl, C.M., D. le Maire and J.R. Munch (2012). "Wage Dispersion and Decentralization of Wage Bargaining", *Journal of Labor Economics*, forthcoming.
- [9] Danish Economic Council (2011). "Danish Economy, Fall 2011." Rosendahls-Schultz: Copenhagen.
- [10] Eissa, N., H.J. Kleven and Claus T. Kreiner (2008). "Evaluation of Four Tax Reforms in the United States: Labor Supply and Welfare Effects for Single Mothers." *Journal of Public Economics* 92, 795-816.
- [11] Feldstein, M. (1995). "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy* 103, 551-572.
- [12] Feldstein, M. (1999). "Tax Avoidance and the Deadweight Loss of the Income Tax." *Review of Economics and Statistics* 81, 674–80.
- [13] Feldstein, M. and M.V. Wrobel (1998). "Can state taxes redistribute income?" *Journal of Public Economics* 68, 369-396.
- [14] Gentry, W. M. and R. G. Hubbard, 2004. "The Effects of Progressive Income Taxation on Job Turnover." *Journal of Public Economics* 88, 2301-2322.

- [15] Hayek, F. A., 1945. "The Use of Knowledge in Society." *American Economic Review* 35, 521-530.
- [16] Immervoll, H., H.J. Kleven, C.T. Kreiner and E. Saez (2007). "Welfare Reform in European Countries: A Micro-Simulation Analysis". *Economic Journal* 117, 1-44.
- [17] Kleven, H.J. and C.T. Kreiner (2005). "Labor Supply Behavior and the Design of Tax and Transfer Policy". *Danish Journal of Economics* 143, 2005, 321-358.
- [18] Kleven, H. J. and C. T. Kreiner (2006). "Beskatning af Arbejdsindkomst i Danmark." Chapter 7 in Torben Tranæs (ed.), *Skat, Arbejde og Lighed*, The Rockwool Foundation, Gyldendal, Copenhagen.
- [19] Kleven, H.J. M. Knudsen, C.T. Kreiner, S. Pedersen and E. Saez (2011). "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica* 79, 651–692.
- [20] Kleven, H.J., C. Landais, E. Saez and E.A. Schultz (2011). "Taxation and International Migration of Top Earners: Evidence from the Foreigner Tax Scheme in Denmark." Preliminary Working Paper, November 2011.
- [21] Kleven, H.J. and E. Schultz (2012). "Estimating Taxable Income Responses using Danish Tax Reforms." Working Paper. London School of Economics.
- [22] Rosholm, M. and M. Svarer. "Endogenous Wage Dispersion in a search-matching Model", *Labour Economics* 11, 623-645.
- [23] Saez, E. (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.
- [24] Saez, E., J. Slemrod, and S.H. Giertz (2012). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50(1), 3–50.
- [25] Slemrod, J. and W. Kopczuk (2002). "The Optimal Elasticity of Taxable Income." *Journal of Public Economics* 84, 91-112.
- [26] Spencer, B. and H. Selin (2012). "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule." Working Paper, Uppsala University, February 2012.
- [27] Topel, R.H. and M.P. Ward (1992). "Job Mobility and the Careers of Young Men." *Quarterly Journal of Economics*, 439-479.