DREAM

The Differences-in-Differences Approach with overlapping differences

- Experimental Verification of Estimation Bias

Hans Bækgaard

Danish Rational Economic Agents Model, DREAM

DREAM Working Paper 2014:3

May 2014

Abstract

This paper demonstrates a hitherto overlooked problem with the differences-in-differences approach where it is applied to panel data and with overlapping differences. A differences-in-differences model with three-year overlapping differences has frequently been applied to measure the behavioural effects of changes to taxation. The problem is that an inherent technical flaw means the differences-in-differences model with overlapping differences seriously underestimate tax elasticities. It is demonstrated that the severity of the problem will depend on the actual situation – such as the nature of the tax changes – but under reasonable and substantiated circumstances the long term tax elasticity is underestimated by more than half. It is, however, not possible to assess the magnitude of the underestimation with any certainty and it is therefore recommended that the method is unreliable and should not be used. In contrast, the error correction model is an alternative dynamic specification that does not suffer from the flaw of the differences-in-differences model with overlapping differences. With explicit distinction between short and long run effects, the error correction specification is recommended as an alternative method for estimating behavioural effects of tax changes.

DREAM, Danish Rational Economic Agents Model. Amaliegade 44, 1256 København K

www.dreammodel.dk

The Differences-in-Differences Approach with overlapping differences

Experimental Verification of Estimation Bias

Hans Bækgaard
hab@dreammodel.dk
+45 33 44 58 93

Introduction

The behavioural effects of changes to taxation are difficult to measure – especially when adjustments are slow and the full effects take years to materialise. Recent research suggests that adjustments are sluggish and there are indications that behavioural adjustments are slower for small than for large tax changes (Chetty et all, 2011). Consequentially, the adjustment process should be represented explicitly so as to avoid the risk of obtaining misleading results.

Since the seminal study by Feldstein (1995), the most widely used approach to measuring the taxable income response to tax changes has been the so-called differences-in-differences approach whereby the change in taxable income is related to the change in key tax parameters for individual taxpayers. Typically, three-year differences have been used in an attempt to measure long-term effects: by taking three-year differences, it is implicitly assumed that the adjustment process has been completed after three years. Identification is achieved by applying the method to a period with sufficiently diverse tax changes across income levels – often spanning several years and with overlapping differences.

In the present paper attention is drawn to a hitherto overlooked problem with particular applications of the differences-in-differences approach that is, where the approach is applied to panel data and with overlapping differences. The problem is essentially that unless the behavioural adjustments are completed already after one year, overlapping differences distort the relationship between the timing of tax changes and the timing of behavioural response. It is demonstrated by example and experimental design, that overlapping differences can – and will most likely – lead to serious underestimation of the tax parameters. Based on substantiated assumptions regarding length and speed of the behavioural response phase-in, the tax elasticity is shown to be underestimated by 25 per cent and up to 60 per cent depending on the precise assumptions. In other words, if the tax elasticity is estimated at 0.1 by the difference-in-difference approach, the 'true' long-term elasticity is more likely to be 0.15 to 0.25. Adding to that, overlapping differences inevitably lead to serious error-term autocorrelation that, if left unattended, may lead to inference bias.

The experimental results are supported by a recent study based on Danish taxation data and a comparison of the differences-in-differences approach to an alternative dynamic specification, a so-called error-correction model, which explicitly distinguishes between short- and long-term effects (Bækgaard, 2012)[1]. The main results suggest that the error-correction model leads to smaller short-term and larger long-run elasticities than the differences-in-differences model, thereby supporting the notion that the differences-in-differences model underestimates the long-term elasticities.

The Differences-in-Differences model (DiD)

The following differences-in-differences model is an example of the problem in question, *cf. Bækgaard (2012)*:

$$(1)\quad \Delta_d \log(y_{it}) = \gamma \log(y_{it-d}) + \xi\Delta_d \log(1-\tau_{it}) + \eta\Delta_d \log(R_{it}) + X_{it}\beta + \varepsilon_{it}$$

y_{it} is taxable income. τ_{it} is the marginal tax rate and R_{it} is virtual income. Lagged income (y_{it-d}) has been added to account for *mean income reversion*. Background variables, X_{it}, control for external factors that may have influenced income changes.

This specification relates the *change* in log-incomes across *d* years to the equivalent *change* in the logarithm of the tax parameters τ_{it} and R_{it}. Typically, three-year differences are used in order to capture long-term effects. The problem is that the strategy only works under specific circumstances. Certainly, if behavioral adjustments take more than one year, a panel model with overlapping differences will inevitably underestimate the elasticities – the question is by how much.[2]

The mechanics of the problem will be demonstrated and a range of experimental designs will assist in establishing the likely magnitude of the problem.

An illustrative example

A simple illustration shows how a differences-in-differences model with overlapping three-year differences can lead to underestimated elasticities. Assume that a tax change is introduced in the fourth year of a panel of eight years (i.e. 5 three-year differences). Further assume that behavioural adjustments to the tax change take place over a three year period with 50 per cent occurring in the year of the tax change and 30 percent in the second year and 20 per cent in the third year. The timing of measured tax changes and behavioural adjustments are shown in the following table.

Table 1 shows that a one-to-one correspondence between measures of tax changes and the behavioural response across three years only exists in one of the five three year periods. The first three three-year differences (year 1 to 4, year 2 to 5 and year 3 to 6) encompass the full effect of the tax change in year 4, whereas taxes are unchanged in the last two three-year differences (year 4 to 7 and year 5 to 8). In contrast, the timing of the

---

[1] An earlier version of Bækgaard (2012) in English (Bækgaard, 2010) is available from http://web.econ.ku.dk/epru/EPRN%20konference%202010/Hans%20B%C3%A6kgaard.01042010_Earned_Income_Response_to_Tax_Changes_in_Denmark.pdf.

[2] It is further noted, that with $d>1$, the error term will invariably be auto-correlated. This may not in itself lead to biased parameters, but it may constitute an inference problem if the autocorrelation is unaccounted for by the estimation technique (most previous studies apply 2SLS, which ignores autocorrelation).

behavioural response is such that only 50 per cent of the full effect falls in the first three-year period and 80 per cent in the second period. In this example, it is only in the third three-year period (year 3 to year 6) that the behavioural effect is 100 per cent. The following two three-year periods only account for 50 and 20 per cent of the income change.

Table 1 Timing of tax and income changes for overlapping three-year periods

Difference period	Share of tax change	Share of income change
Year 1 to year 4	100 per cent	50 per cent
Year 2 to year 5	100 per cent	80 per cent.
Year 3 to year 6	100 per cent	100 per cent
Year 4 to year 7	0 per cent	50 per cent
Year 5 to year 8	0 per cent	20 per cent

A DiD model estimated on the basis of the situation displayed in the above table will underestimate the behavioural parameters due to the lack of correspondence between effect and response.[3]

We will now turn to experimentation in order to assess the magnitude of the problem.

Experimental evidence

Without loss of generality, the experiments will apply the following simplified version of the DiD model (1):

(2) $$\Delta_3 \log(y_{it}) = \alpha + \gamma \log(y_{it-3}) + \xi \Delta_3 \log(1 - \tau_{it}) + \varepsilon_{it}$$

This model is estimated on the basis of simulated data generated from assumptions about the underlying relationship between tax changes – represented by the change in marginal net income $(1 - \tau_{it})$ – and the change in taxable income y_{it}.

The data generating process is described by the following steps:

1. Create a sample of taxable incomes in year $t = 1$, $y_{i1}, i = 1,\ldots,N$ such that the sample is grouped in $j = 1,\ldots,J$ subgroups
2. Define subgroup specific marginal tax rates for the entire panel period i.e. τ_{jt} $j = 1,\ldots,J$, $t = 1,\ldots,T$

[3] It is obvious that a one-to-one correspondence only exists for the year 3 to year 6 differences, and applying DiD to only this period will reveal unbiased behavioural parameters. This is, indeed, a standard way of applying the one-period DiD model to a specific tax reform; an approach that will work assuming that other conditions are fulfilled (e.g. full phase-in after three years, no tax changes in year 5 and 6, no income shifting between years, etc.).

3. Define the timing of the phase-in of the behavioural response i.e. $b_t, t = s,...,s+B$ where $b_t \sum_{t=s}^{s+B} b_t = 1$ is the share of the full response that occur in year t, s is the year of the tax change and B the number of years until full phase-in is complete.
4. Generate $y_{it}, i = 1,...,N$ recursively for $t = 2,...T$ by drawing $\varepsilon_{it} \sim n(0,\sigma^2)$ and applying the equation:

(3) $$\log(y_{it}) = \alpha^1 + \gamma^1 \log(y_{it-1}) + \xi \sum_{s=0}^{B-1} b_{t-s} \Delta(1-\tau_{it-s}) + \varepsilon_{it}$$

There are two important things to note in relation to this experimental design. Firstly, the taxation implied by the marginal tax rates τ_{jt} ($j = 1,...,J$, $t = 1,...,T$) is predetermined in the sense that individuals do not change tax brackets when their income changes as the income generating process proceeds recursively. While this is clearly not a realistic description of the tax system, it does mean that the endogeneity problem caused by movements between tax brackets is avoided.[4] Secondly, if $\gamma^1 = 1$ the income generating process does not imply mean income reversion and, by implication, there is no need to include lagged income in (2). In contrast, if $\gamma^1 < 1$ there will be mean reversion and as a result $\gamma < 0$.[5] In the following, we will therefore test different assumptions about γ^1.

The absence of endogeneity and (optionally) mean income reversion is a result of a simplified income process. It does not, however, have any implications for the phenomenon that we are seeking to illustrate, namely how the mathematics of overlapping differences combined with sluggish behavioural response tends to lead to underestimated elasticities.

The main experiment is based on the marginal tax rates in the Danish tax system during the period 1994 to 2006, *cf. Figure 1*.[6] The sample of initial taxable incomes in 1994 (step 1 above) have been drawn for 500,000 individuals[7] from five income groups centered in the five equally spaced 1994-income levels in the figure, that represent the four main tax brackets:

120,000 DKK: in the bottom tax bracket during the whole period

160,000 DKK: in the middle tax bracket from 1994 to 2000 and the bottom tax bracket from 2001 to 2006

---

[4] The issues relating to endogeneity and mean income reversion are both discussed intensively in the literature, e.g. Bækgaard (2010).

[5] It is noted that Bækgaard (2013) estimated γ-values in (1) between -0.448 and -0.329 – with different values for males and females and education groups. These values correspond to γ^1-values in the range 0.820 to 0.876.

[6] The period 1994 to 2006 coincides with the period of examination in Bækgaard (2010 and 2012), which provides a description of the tax changes during this period.

[7] The sample of 2,500,000 individuals is roughly equivalent to the size of the Danish workforce.

200,000 DKK: in the '6 per cent'[8] tax bracket from 1994 to 1995, the middle tax bracket from 1996 2003 and the bottom tax bracket from 2004 to 2006

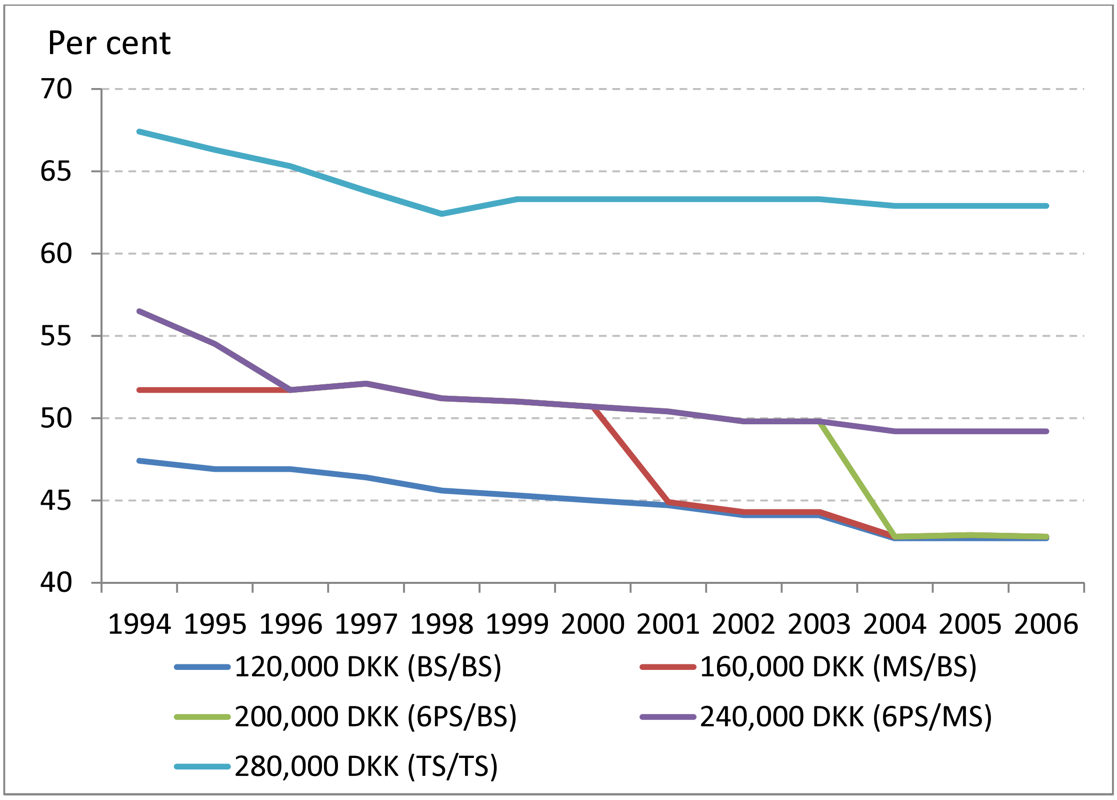
240,000 DKK: in the '6 per cent' tax bracket from 1994 to 1995 and the middle tax bracket from 1996 to 2006

280,000 DKK: in the top tax bracket during the whole period

A noise term drawn from $n(0,\sigma_y^2)$ has been added to the initial values (σ_y=10,000 has been used – the magnitude of the noise does not influence results, only the measurement error). With no loss of generality, all the simulations were based on a long-term (i.e. fully phased-in) tax elasticity of ξ=0.2.

The first experiments are based on the actual development in Danish marginal tax rates over the period from 1994 to 2006 for the above five income groups defined by the four main tax brackets, *cf. Figure 1*. Although all income groups have faced reduced marginal tax rates over the period, the relative size and the timing of the reductions are sufficiently diverse to allow identification of the tax parameters.

Figure 1 Marginal tax rates for selected income levels (1994-level), 1994 – 2006



Soure: Bækgaard (2010).

The following table shows regression results for (3) under different assumptions regarding the phase-in pattern of behavioural response $b_t, t = s,...,s+B$ and the γ^1 -

[8] The so-called '6 per cent' tax was phased out from 1994 to 1996.

parameter. Recall that the latter introduces mean income reversion when γ^1<1. The first thing to note is how the elasticity parameter depends on the assumption about the phase-in pattern of behavioural adjustment. As predicted by the illustrative example (see above), the estimated elasticity declines as the phase-in is prolonged. Indeed, the elasticity declines from 0.205 (~0.2) to 0.160 with a phase-in over three years (0.5 in year 1 and 0.25 in year 2 and 3) and further to 0.132 with a five year phase-in (0.4 in year 1, 0.25 in year 2, 0.2 in year 3, 0.1 in year 4 and 0.05 in year 5), and it declines to 0.109 when the phase-in process is slowed further (0.3 in year 1, 0.25 in year 2, 0.2 in year 3, 0.15 in year 4 and 0.1 in year 5). This is a direct result of the in-built bias of the DiD model with overlapping differences. The estimation bias is increased slightly by introducing mean income reversion by including lagged income.

Table 2 Experiments with the Differences-in-Differences model with three-year overlapping differences: actual marginal tax rates 1994 to 2006

Assumptions		Model		
Phase-in Year 1 to 5	α^1 / γ^1	**Without γ** ξ	**With γ** ξ	γ
1/0/0/0/0	-0.13 / 1.0	0.205	0.205	0.0001
0.5/0.25/0.25/0/0	-0.13 / 1.0	0.160	0.156	0.0002
0.4/0.25/0.2/0.1/0.05	-0.13 / 1.0	0.132	0.126	0.0003
0.3/0.25/0.2/0.15/0.1	-0.13 / 1.0	0.109	0.101	0.0004
1/0/0/0/0	1.13 / 0.9	0.039	0.183	-0.271
0.5/0.25/0.25/0/0	1.13 / 0.9	0.003	0.144	-0.271
0.4/0.25/0.2/0.1/0.05	1.13 / 0.9	-0.023	0.117	-0.271
0.3/0.25/0.2/0.15/0.1	1.13 / 0.9	-0.044	0.094	-0.271
Effect of the γ^1 -parameter				
1/0/0/0/0	-0.13 / 1.0	0.205	0.205	0.000
1/0/0/0/0	1.13 / 0.9	0.039	0.183	-0.271
1/0/0/0/0	2.35 / 0.8	-0.226	0.163	-0.488
1/0/0/0/0	3.60 / 0.7	-0.467	0.145	-0.657
0.4/0.25/0.2/0.1/0.05	-0.13 / 1.0	0.132	0.126	0.000
0.4/0.25/0.2/0.1/0.05	1.13 / 0.9	-0.023	0.117	-0.271
0.4/0.25/0.2/0.1/0.05	2.35 / 0.8	-0.276	0.108	-0.488
0.4/0.25/0.2/0.1/0.05	3.60 / 0.7	-0.504	0.099	-0.657

Notes: With experimental data, the parameter standard errors are fairly constant: st.dev.(ξ)≈0.004; st.dev.(γ)≈0.0001. All experiments are run with 500,000 individuals in each of the five groups in each year of the 10 three year differences in the period 1994 to 2006. The α^1 -parameters have been chosen to keep average incomes roughly constant from year to year, but otherwise make no difference to the results. The 'Phase-in year 1 to 5' shows the share of the full effect in year 1 to 5, e.g. '1/0/0/0/0' means full effect in the first year of implementation. The estimated γ -values correspond closely to the experimental γ^1 -values and the transformation from one to three lags and (2) being in differences: $\gamma = -(1-\gamma^{1^3})$.

With the γ^1 -parameter set at 0.9, meaningful elasticities are, as expected, achieved only when lagged income is also included in the estimation. Nevertheless, even after accounting for the mean income reversion, the elasticities are further biased downward by around 10 per cent.

The second part of Table 2 illustrates the effect of the γ^1 -parameter. The results show, as expected, that the tax parameter turns negative as γ^1 is reduced. That is, unless mean reversion is accounted for by including lagged income in the equation, in which case the estimated elasticity is positive albeit downward bias. Reducing γ^1 increases the negative bias.

The next experiment tests the importance of the nature of the tax changes during the period under scrutiny. To illustrate, two hypothetical series of marginal tax rates for the five income groups have been invented for the 13-year period:

1. A one-off reduction in the marginal tax rate for four income groups (in different years) and constant rates for one income group, *cf. Figure 2*.
2. A one-off increase in the marginal tax rate for four income groups (in different years) and constant rates for one income group, *cf. Figure 3*.

Table 3 Experiments with the Differences-in-Differences model with three-year overlapping differences: importance of nature of tax changes

Assumptions		Marginal tax rate series		
Phase-in Year 1 to 5	α^1/γ^1	Actual MTRs ξ	One-off decrease ξ	One-off increase ξ
1/0/0/0/0	-0.13 / 1.0	0.205	0.202	0.198
0.5/0.25/0.25/0/0	-0.13 / 1.0	0.160	0.135	0.138
0.4/0.25/0.2/0.1/0.05	-0.13 / 1.0	0.132	0.108	0.111
0.3/0.25/0.2/0.15/0.1	-0.13 / 1.0	0.109	0.085	0.089
1/0/0/0/0	1.13 / 0.9	0.183	0.188	0.179
0.5/0.25/0.25/0/0	1.13 / 0.9	0.144	0.121	0.124
0.4/0.25/0.2/0.1/0.05	1.13 / 0.9	0.117	0.097	0.099
0.3/0.25/0.2/0.15/0.1	1.13 / 0.9	0.094	0.075	0.078

Notes: See Table 2.

Figure 2 Experimental marginal tax rates: one-off decreases, 1994 – 2006

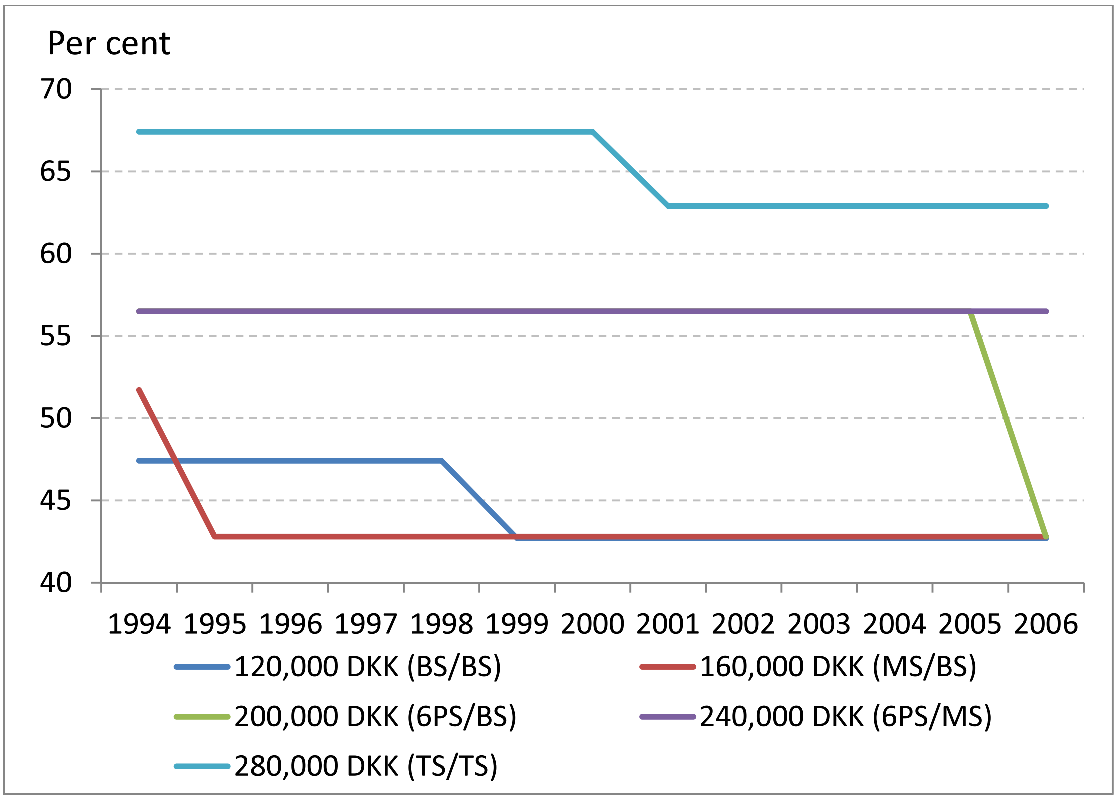
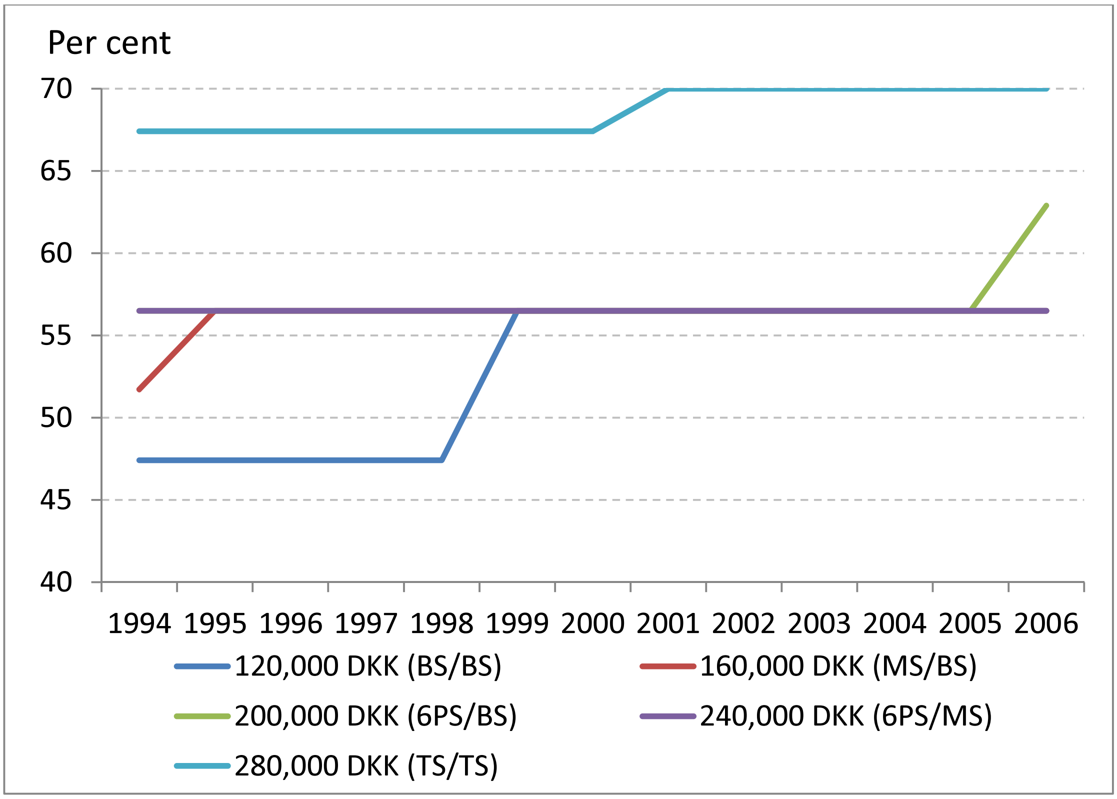


Figure 3 Experimental marginal tax rates: one-off increases, 1994 – 2006



The estimation results for these marginal tax rate series are compared with the results for the actual tax rates for different assumptions regarding behavioural response dynamics and for γ^1 set to 1 and 0.9 respectively, *cf. Table 3*. The results suggest that the nature of the tax changes matter somewhat, but only when behavioural adjustments are sluggish. The bias is worse when the tax changes only occur once instead of the more gradual change in actual MTRs over the period *cf. Figure 1*. In contrast, it does not appear to matter whether marginal tax rates go up or down.

Discussion

The experimental results confirm that, if the behavioural adjustment process continues beyond the year of implementation, the differences-in-differences model with overlapping differences systematically underestimate the 'true' elasticities – the slower the adjustment process, the stronger the negative bias. The experiments with different behavioural phase-in patterns show that a five-year phase-in (with 0.3 in year 1, 0.25 in year 2, 0.2 in year 3, 0.15 in year 4 and 0.1 in year 5) leads to a negative bias of around 50 per cent based on actual tax changes over the analyzed period (1994 to 2006). The bias is even stronger, over 60 per cent, when based on one-off tax changes.

While it is not possible to quantify the bias accurately on the basis of experimental data alone, an estimate of the magnitude of the problem is obtained by looking at the results from a recent study that compares the results obtained by a DiD model with overlapping differences and those from an error correction model (ECM), *cf. Bækgaard (2010 and 2012)*. The ECM is an alternative approach that explicitly distinguishes between short- and long-run elasticities – it does not suffer from the specification problems of the DiD, *see Box 1*. Separate results for males and females are shown in *Table 4*.

Box 1 The Error Correction Model (ECM)

The error correction model provides an alternative to the DiD approach. The ECM identifies short and long term effects separately by the following specification:

$$\Delta \log(y_{it}) = \xi_s \Delta(1-\tau_{it}) + \eta_s \Delta R_{it} + \Delta X_{it}^1 \beta_1 +$$

$$(\gamma - 1)[\log(y_{it-1}) - (\alpha_i + \xi_l (1-\tau_{it-1}) + \eta_l R_{it-1} + X_{it-1}^2 \beta_2)] + \varepsilon_{it}$$

ξ_s and ξ_l, and η_s and η_l are the short and long run substitution and income elasticities respectively.

The first thing to note is that the ECM model's long-run elasticities are considerably larger than the short-run (1st year) elasticities. For males the short-run substitution elasticity is estimated at 0.073 and the long-run elasticity at 0.265 or 0.388 without income effect – either way, the 1st year effect is less than the 30 per cent of the full effect used in the 'slow' phase-in in the above experiments. Interestingly, the substitution elasticity estimated by the DiD is 0.109, which amounts to around 40 per cent of the long-run elasticity. The results for males thus lend support to a DID downward bias around 60 per cent.

The results for females are more ambiguous due to problems with simultaneous identification of substitution and income elasticities, *cf. Table 4* (see also Bækgaard 2010 and 2012). However, the ECM-results for females without income effects point to almost identical relationships as for males in terms of the relativities between the short-run ECM and the DiD elasticities on the on hand side and the ECM long run elastcities on the other. The short-run elasticity at 0.043 is around 30 per cent of the long-run elasticity at 0.132 and the DiD estimate at 0.056 is around 40 per cent of the long-run effect – i.e. the DiD underestimates the long-run elasticity by 60 per cent.

Table 4 Estimation results for the differences-in-differences and the error correction models

	Men			Women		
	ECM Income effect	ECM No Income effect[1]	DiD	ECM Income effect	ECM No Income effect[1]	DiD
Substitution elasticity (ξ)						
- short run	0.073	0.076	-	0.014	0.043	-
	(0.023)	(0.020)		(0.021)	(0.021)	
- long run	0.265	0.388	0.109	0.023	0.132	0.056
	(0.047)	(0.046)	(0.013)	(0.041)	(0.043)	(0.012)
Income elasticity (η)						
- short run	-0.001	-	-	-0.012	-	-
	(0.002)			(0.003)		
- long run	-0.042	-	-0.017	-0.037	-	-0.013
	(0.003)		(0.002)	(0.005)		(0.002)

[1] Virtual income is excluded from the estimation.

Source: Bækgaard (2010).

The story told by the estimated income elasticities obtained from the DiD and the error correction models also lend support to the claim that the DiD model seriously underestimates the long term effects, *cf. Table 4*. Indeed, the estimated income elasticities from the DiD model are only around one third of the long term effects from the ECM model. The stronger negative bias for the income elasticities could be explained by a slower adjustment process for the income effect, at least for males where the short-run income elasticity is insignificant while the long-run elasticity is estimated at -0.042.

A number of studies have used a DiD model with overlapping differences to estimate the elasticity of taxable income. On the basis of the above analysis it seems reasonable to conclude that the elasticities obtained by these studies underestimate the long run income response to tax changes.

Apart for the DiD model in Bækgaard (2010 and 2012), a DID model has been applied by Gruber and Saez (2002), Chetty m.fl. (2011) and more recently for Denmark by Kleven and Schultz (2013). The latter study finds substitution and income elasticities that are broadly in line with the DiD elasticities in Table 4. Their substitution elasticities are estimated at around 0.05 for wage earners and 0.10 for self-employed (joint estimates for males and females) and like the DiD estimates in Table 4, the methodology problems imply that these elasticities seriously underestimate the 'true' long-term elasticites. This notion is supported by the much larger estimate at around 0.2 obtained by Kleven and Schultz (2013) when only one period (1986-89) is used to provide an isolated estimate of the effect of the 1987 tax reform. The authors interpret this as evidence that large tax reforms produce larger effects than small reforms, but the larger estimate is more likely a result of it being obtained by a one-difference estimation, which is not subject to the underestimation bias documented here.

References

Bækgaard, H. 2010. *Earned Income Response to Tax Changes in Denmark*, Working Paper presented at the EPRN conference June 2010, Copenhagen University.

Bækgaard, H. 2012. *Elasticiteten af skattepligtig arbejdsindkomst*. Nationaløkonomisk Tidsskrift 150 (2012), pp. 119-143.

Chetty, R., J. Friedman, T. Olsen og L. Pistaferri 2011. *Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records*, The Quarterly Journal of Economics 126 (2), pp. 749-804.

Feldstein, M. 1995. *The Effect of Marginal Tax Rates on Taxable Income: a Panel Study of the 1986 Tax Reform Act*, The Journal of Political Economy, Vol. 103, No. 3., pp. 551-72.

Gruber, J. og E. Saez. 2002. *The Elasticity of Taxable Income: Evidence and Implications*, Journal of Public Economics 84, pp.1-32.

Kleven, H. og E. A. Schultz (2013) *Estimating Taxable Income Responses using Danish Tax Reforms*, LSE Working Paper, *American Economic Journal: Economic Policy*, forthcoming.