ESTIMATING LABOUR SUPPLY RESPONSES USING PROVINCIAL TAX REFORMS

BENJAMIN MACLEAN SAND

Abstract. This paper examines labour supply effects using variation in after-tax wages from a variety of recent provincial tax reforms in Canada. Prior to 2000, nearly all provincial governments set income tax rates as a percentage of federal income tax. Since then, the federal-provincial tax-on-income (TONI) agreement has allowed provincial governments to set their own rates, credits, and exemption levels. This dissociation of tax schedules, along with several major tax reforms, has resulted in marginal tax rates and thresholds that differ significantly between provinces. The variation induced by these tax reforms occurred both within and across provinces and affected virtually all income groups. This offers a unique opportunity to study labour supply responses for a more general segment of the population than 'natural experiments' normally permit. Future work will exploit this variation using longitudinal data from a confidential version of the Survey of Labour Income and Dynamics (SLID) that allows me to observe the same individuals before and after the tax reforms. Preliminary results (below) using publicly available data typically indicate small labour supply responses that are not significantly different from zero, with larger responses for individuals with younger children. The results are similar for both men and women.

1. Introduction

This paper employs a number of methods to empirically evaluate claims that marginal tax cuts spur economic growth by providing Canadians with incentives to work harder and increase their labour supply. Although these claims are made by the government of Alberta, among others, there is little evidence to date that marginal tax cuts will create the large efficiency gains they assert. Since, in general, tax policy designers face a trade off of efficiency with distributional goals (McKenzie 2000), it is essential to evaluate these arguments for tax reform.

Date: September 2005.

I am grateful to Kevin Milligan for generously allowing to use his Canadian Tax Calculator.

There is a large body of economic literature evaluating labour supply responses to tax policy. The linkage between government tax policy and people's employment and hours of work decisions is of great importance to policy designers. These labour supply responses play a large part in overall economic activity, quality of life, and government income tax revenue. However, in the past there has been little consensus of the effect of tax policy on the elasticity of labour supply, due to complications in its estimation (Eissa 1995) (Blundell, Duncan, and Meghir 1998). For example, individual differences in tastes for work and leisure, that cannot be controlled for using observable characteristics, make the isolation of the effect of marginal tax rates on labour supply difficult. In addition, since classical economic theory leads to mostly ambiguous predictions of labour responses, the empirical estimation of such effects is of paramount interest to both labour econometricians and policy makers.

This empirical research is particularly important for Canada, since much of the past research has focused on the U.S. and the U.K. or on small and mostly disadvantaged subgroups of the Canadian population (Blundell and MaCurdy 1999). This paper focuses on a more general segment of the Canadian population, using a number of econometric techniques to take advantage of a unique opportunity to study a Canadian tax 'natural experiment'. In particular, the focus of this study is on the labour response to Alberta and British Columbia's 2001 tax reforms, where Alberta abandoned a progressive marginal tax system in favour of a 10 percent single tax rate, while British Columbia eliminated surtaxes and added additional brackets. I use the variation induced by the differing progressivity of each tax reform to identify the parameters of interest. This type of 'natural experiment' analysis has enjoyed wide spread acceptance in econometric literature due to its ability to deal with criticisms of previous labour supply studies, such as non-observable tastes, by using exogenous after tax wage changes to identify labour responses over time (Blundell and MaCurdy 1999).[1]

2. Literature Review

There is a large body of empirical research on the effects of taxation on labour supply producing wide range of estimates. This range of estimates is in part due to differing methodology and assumptions made by researchers to deal with some of the difficulties of estimating a labour supply elasticity. In general, these difficulties arise because of non-linearities in tax schedules, endogeneity, and selection problems (Blundell, Duncan, and Meghir 1998)

[1]It is not, however, without its criticisms. Heckman (1996) criticises the method in certain situations for being too "atheoretical".

(Eissa 1995). This literature is too vast for a rigorous account to be given here and, therefore, I provide only a brief description of some of the more important themes before turning to studies that more closely relate to the approach taken in this proposal.

Earlier research on labour supply attempting to overcome problems of selection and endogeniety of wage (and marginal tax rates) relied on corrections that largely required an instrumental variable or exclusion restriction. Often times these corrections produce non-robust results that depend on the particular (and sometimes, arbitrary) assumptions made on the excluded variable(s). Mroz (1987) gives a comprehensive account of such problems and highlights sensitivity of results to the assumptions made to correct for them. Sensitivity concerns also arise due to the differing approaches used to deal with the non-linear nature of tax schedules. Earlier estimation procedures that were functional-form dependent produced unreliable parameter estimates (Blundell and MaCurdy 1999). MaCurdy, Green, and Paarch (1990) find that much of the disparity of parameter estimates between studies can be explained by the differing methods used to deal with non-linear budget constraints that impose implicit parameter restrictions at kink points.

As a result of these problems, some researchers have increasingly relied on "natural experiment" methods to directly evaluate tax policy on labour supply decisions (Blundell and MaCurdy 1999). These methods exploit policy induced *exogenous* change in after tax wages and income and are not as dependent on arbitrary exclusion restrictions or on function form for identification (Eissa 1995). In what follows, I discuss recent empirical studies that use a natural experiment framework.

Eissa (1995) uses a natural experiment approach to estimate the effects of the U.S. Tax Reform Act (TRA) of 1986 on the labour supply of married women. Her strategy uses repeated cross sections from the March Current Population Survey (CPS) from 1984 to 1986 and from 1990 to 1992, to capture the pre and post tax reform behaviour of the women in her sample. The TRA had a different impact on married women along the income distribution[2]. In particular, the women at or above the 99th percentile of the income distribution received marginal tax reductions in the order of 30 percent. Eissa selects these women as her treatment group and uses two control groups: married women at the 75th and 90th percentiles of the distribution. These groups had far smaller reductions in the marginal tax rates.

[2] Eissa uses spousal income and joint capital income to define a woman's position in the income distribution.

Eissa's identification strategy uses a difference-in-difference estimator to compare the working behaviour of the treatment and the control groups before and after the tax reform. Her dependent variables include participation and annual hours worked. She adds to the simple difference-in-difference estimator a regression method to control for the groups different observable characteristics and participation into the labour force. Her results for her preferred control group suggest that the elasticity of total labour supply with respect to after-tax wage for high income married women is between 0.6 and 1.0. Her estimated elasticity with respect to participation are smaller and are between 0.2 and 0.4. When Eissa included in one of her specifications an education-treatment interaction, it reduced the estimated participation elasticity. This suggests that part of the increase in participation is from women that are more highly educated, which may have a possible demand side interpretation.

Eissa's methodology, however, has not gone without its criticisms. The difference-in-differences approach taken in her study relies on strongly restrictive structural assumptions. In particular, it requires that the control and treatment groups do not change in a non-random way, and that common time trends operate on both groups. Her identification strategy has been criticised on both accounts. In the first case, it is plausible that women could have changed groups as a result of the tax reform either because of capital income tax elasticity or endogeneity of the husband's income (Heckman 1996). Also, the common trends assumption may be violated due to the widening wage distribution between all groups during the period of her study (Blundell, Duncan, and Meghir 1998)[3].

Blundell, Duncan, and Meghir (1998) use a series of tax reforms in the U.K. over the 1980s to study the labour supply of married or cohabiting women. They use a data from the U.K. Family Expenditure Survey over the period of 1978 to 1992, and employ a grouped Wald estimator to estimate substitution and income effects of labour supply to changes in after tax income. Also estimated was a difference-in-difference estimator. In addition to offering evidence on the labour supply of married women in the U.K., the paper also offers a critic of the group-by-tax status method pursued by Eissa. Their "natural experiment" framework groups married women based on cohort and education level. Tax policy reforms over the period of interest affected these groups differently due to the progressivity of the tax system.

[3]Heckman (1996) condemns the approach on other grounds as well. For example, he concludes that the approach is too "athoeretical", does not use all identifying information, and does not produce results that are economically interpretable or comparable to other studies. His assessment is that "the only thing going for the method ... is computational convenience" [p.33].

This study was more ambitious than Eissa's in that they estimated both a substitution effect and an income effect, not just a total labour supply elasticity. To do this, they use a model consistent with life-cycle labour supply. This specification includes after tax wages and other household income. Since these variables are endogenous, Blundell et. al. instrument with group-time interactions. This method exploited the demand side variation in wages between the groups over the period of interest, which was assumed exogenous to the hours equation. The essence of this method is that it compares the after tax wages and other income between groups who were affected differently by tax policy. Again, this approach is strongly dependent on the standard assumptions of a Wald estimator. In particular, grouping could not be endogenous to tax reforms and average differences in labour supply, given characteristics, between groups has to be constant over time.

Their findings indicate that the uncompensated wage elasticities of married women are relatively small but positive. For women with no children the reported elasticity was 0.14 and it increased somewhat for women with younger children. They also report their findings when the analysis is performed on groups defined by tax status. They find that the results differ from their preferred IV method, and trace the discrepancy back to the change in composition of the tax-payer group over time. They conclude that tax reform should take into consideration the behavioural response of workers, since they find that taxation has efficiency costs in terms of reduced labour supply.

3. The Tax Reforms

In Canada, personal income taxes are imposed by both the federal and the provincial governments. Until the year 2000, provincial governments set taxes as a percentage of federal taxes (except Quebec). However, by 2001 all provincial governments calculated taxes based on federally defined taxable income rather than federal taxes. This change in policy allows provincial governments to pursue more freely their own goals of redistribution and efficiency.

3.0.1. *Alberta.* In January, 2001, Alberta introduced Canada's first single-rate personal income tax system, which represents a significant departure from the progressive marginal tax rates found elsewhere in Canada and other countries. This tax structure reform was accompanied by a significant tax reduction, elimination of the high-income surtax, and the introduction of the largest personal and spousal exemptions in the country. This shift of tax structure had large distributional consequences, shifting tax burden away from particularly low and very high income earners (McMillan 2000).

Prior to 2001, Alberta's income tax system was a percentage of the federal taxes paid. This amount was 44% in the year 2000. In addition to the marginal tax rate faced by individuals, there were also surtaxes to pay on top of the marginal rates. In 1999, the top federal and provincial combined rate was about 44.7%. This was reduced to about 39% following the 2001 tax reform. For some tax payers, particularly in the middle of the income distribution, the tax reduction was far less. However, for very low income earners, larger provincial exemptions resulted in large reductions in federal/provincial combined rates.

3.0.2. *British Columbia.* British Columbia employed the tax-on-income system in 2000, but changed its tax rates relatively little from the tax-on-tax method used in 1999. Significant reform in the British Columbian tax system did not occur until the left leaning NDP lost a provincial election to the centre-right Liberals in May of 2001. Only weeks later, the Liberals cut taxes significantly and increased the number of tax brackets from three to five. These reductions stemmed almost entirely from the elimination of provincial surtaxes, benefiting high income earners relatively more than low. The reform took place on June 6th, although the pay-as-you-go tax collection was adjusted for the remainder of the year to reflect the fact that the changes where to be effective for the entire year - beginning January 1st, 2001. The following year additional reductions were made. The reforms resulted in the lowest marginal tax rates in the country for those at the bottom of the distribution and a top rate that is was reduced to a level second only to Alberta.

3.0.3. *Federal Government.* In addition to provincial tax reforms, the federal government also altered marginal tax rates and brackets in 2001. This reform increased the number of tax brackets from three to four, and was accompanied by small reductions to rates for earners in the lower and middle part of the income distribution.

4. Empirical Approach

To examine the effects of the provincial tax reforms in British Columbia and Alberta, the key evaluation problem is to construct the counterfactual outcomes to represent what would have happened in their absence. A comparison of the Albertan outcomes under the tax reform and the estimated counterfactual is the estimated effect of the program. As a first pass at this problem, this paper uses difference-in-differences (DD) techniques to estimate the effect of the tax reform before moving onto a more parametrised model that more fully exploits the variation in the data induced by the provincial reforms. Both methods rely on the longitudinal nature of the data. That is, since we observe outcomes before and after the policy change, we can use the change in the outcome variable along with a control group to identify the treatment effects.

The main concern of this paper is to examine the tax reforms effect on labour supply. I focus on two margins of supply: participation and hours worked per week conditional on participation. Participation is defined as employment, and I create a binary variable equal to 1 if an individual is employed (including self employment) and 0 otherwise. Analysis on hours worked per week is constrained to those workers who are employees.

4.1. **Difference-in-differences.** The most straightforward way to examine the impact of provincial tax reform is the difference-in-difference (DD) framework. This procedure simply compares the difference in the before/after outcomes of the treatment group (a province with a reform) to a control group (another province). Under the assumption that the evolution of the outcome variable would have been the same between the two groups in the absence of the tax reform, this will produce consistent estimates of the impact of the tax reform on the treated province.

To be more formal, define the treatment as living in a reform province in the post 2001 period. Thus, the treatment is being exposed to a 2001 tax reform. Let the potential outcome conditional on treatment by Y_1 and Y_0 conditional on no treatment. We group individuals based on provincial residence. Denote the treatment province by (T), and the control group is another Canadian province (C). Let $t = 1$ denote the years after the tax reform and $t = 0$ the years before. The parameter of interest throughout this analysis is the average treatment effect on the treated. This is one of the more common parameters in the the programme evaluation literature (Todd and Smith 2003) (Heckman, Ichimura, and Todd 1997). Define the parameter of interest by

$$\text{ATET} = \mathbb{E}\left[Y_1 - Y_0 | X, T, t = 1\right] = \mathbb{E}\left[Y_1 | X, T, t = 1\right] - \mathbb{E}\left[Y_0 | X, T, t = 1\right].$$

Since we cannot observe the outcomes Y_1 and Y_0 for anyone, we have a missing data problem. In particular, to estimate the parameter ATET the counterfactual mean outcome $\mathbb{E}\left[Y_0 | X, T, t = 1\right]$ must be estimated.

To estimate this counterfactual, the key assumption made in this paper is[4]

Assumption 1.

$$\mathbb{E}\left[Y_{0it}|X_{it},T,t=1\right]-\mathbb{E}\left[Y_{0it}|,X_{it},T,t=0\right]=\mathbb{E}\left[Y_{0it'}|X_{it},C,t=1\right]-\mathbb{E}\left[Y_{0it}|X_{it},C,t=0\right]$$

This assumption states that in the absence of the tax reform, growth in the outcome variable for a reform province would be the same as in the control province. The assumption implies common time effects operate on both the control and treatment group, and the composition of the groups do not change in a non-random way in response to the tax reform. Thus, the control group used is crucial for this analysis. This is discussed in more detail later on. We can solve for our missing counterfactual

$$\mathbb{E}\left[Y_{0it}|X_{it},T,t=1\right]=\mathbb{E}\left[Y_{0it}|,X_{it},T,t=0\right]+\mathbb{E}\left[Y_{0it'}|X_{it'},C,t=1\right]-\mathbb{E}\left[Y_{0it}|X_{it},C,t=0\right].$$

This procedure is easily implemented in a regression framework. Consider the regression equation,

$$Y_{it}=\beta_0+\beta_1X_{it}+\delta_iD_{it}+u_{it}$$

where i and t index individuals and time, respectively. X_{it} is a vector if individual observable characteristics and D_{it} is the treatment indicator that equals 1 if individual i is living in a reform province in time t. The error term is represented by u_{it}. To incorporate the above assumption, we can decompose its expectation into province and time fixed effects: $\mathbb{E}\left[u_{it}|p,t\right]=a_p+m_t$, where a_p is a provincial fixed effect and m_t is a time effect that is common across provinces.

This assumption states that unobservable differences in the outcome variable between provinces' can be explained by a time-constant provincial effect and a common time effect. In other words, the growth in the outcome variable would be the same in Alberta and a comparison province in the absence of the tax reform. This is exactly the same assumption made above. Using this assumption the above equation can be estimated by including year and group specific intercepts:

[4]This presentation follows Blundell, Costa Dias, Meghir, and Van Reenen (2003). They point out how common this assumption is in evaluations of this type.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_i D_{it} + a_p + m_t + \epsilon_{it}$$

where ϵ_{it} now represents the transient component of the error term. Indexing δ by i indicates that the effect of the reform can be heterogeneous. Then under this assumption, the DD framework estimates

$$\hat{\delta} = \mathbb{E}\left[\delta_i | D_i = 1\right] = ATET.$$

We can give the estimated $\hat{\delta}$ a causal interpretation if $\mathbb{E}\left[\epsilon_i | a_{ip}, m_{it}, D_{it}, X_{it}\right] = 0$.

Although this approach is widely used in the evaluation of policy, it has several limitations. A major limitation is that this procedure is heavily dependent on it's identifying assumption, that, if violated, calls into question the results ((Blundell and MaCurdy 1999); (Heckman 1996)). However, even if the error components assumption is correct, this approach suffers from at least two major drawbacks. First, although the estimated parameter $\hat{\delta}$ neatly summarises the policy effect, it is not really informative about elasticity responses (Gruber 1996). Secondly, it does not take advantage of all the potentially identifying variation induced by the provincial tax reforms. For example, Alberta's flattening of the tax schedule left tax rates at certain income levels virtually unchanged, while BC's more general lowering of taxes reduced rates at all income levels - particularly at higher incomes. Thus, we can attain further identifying information by exploiting not only the variation in tax changes between provinces but *within* provinces.

4.2. **Reduced form Models.** To take advantage of differential changes in taxes within a province, I need to find an additional control group within a province whose tax incidence's differ. For example, suppose that I could identify skill groups that have the property that the higher one's skill the higher one's income. Since changes in marginal tax rates (MTRs) within provinces differ by income levels, one can compare the changes labour supply between skill groups between periods to identify the effect of MTRs. For example, in an important paper by Blundell, Duncan, and Meghir (1998), married women in the U.K. during the 1980s were grouped by cohort and education level. Since these groups of women where treated differently on average by the tax system during that time, the authors were able to identify wage elasticities by comparing the labour supply responses of these groups. With only one province, this method is just the fixed effects or difference-in-difference method.

The identifying assumption is that the average difference in labour supply between groups be vertically parallel through time in the absence of reforms.

As Blundell, et. al. (1998) point out, this type of fixed-effects estimator is equivalent to the Wald or Instrumental Variables (IV) estimator with the excluded instruments the time-group interactions. Thinking about the problem in this way helps to clarify the potential sources of identifying variation caused by provincial tax reforms and the assumptions need for their use to be valid. For example, suppose that in period one ($t = 1$) that provinces implement tax reforms and that the reforms differ by skill group and province. Then there are three potentially useful sources of variation of MTRs induced by these reforms: (1) Province/time (2) skill group/time (3) province/skill group/time. Suppose we have a sample of workers of various skills from different provinces over at least two time periods and have the reduced form equation of some labour supply outcome Y_{it}[5].:

$$Y_{it} = c + \delta_i MTR_{it} + u_{it}$$

where Y_{it} is some labour supply outcome and u_{it} is an error term. The problems in consistently estimating $\hat{\delta}$ include possible correlation between the unobservable error term and MTRs, the presence of common time shocks, provincial specific shocks, skill specific shocks, and various skill or provincial time trends. Which of these are important depends on what our assumptions about the structure of the error term are. We can begin by making the following assumption:

Assumption 2. ***triple-DD***

$$\mathbb{E}\left[u_{it}|P_i, T_i, S_i\right] = c + \beta_0 P_i + \beta_1 S_i + \beta_2 T_i + \beta_3 P_i \cdot S_i + \beta_4 P_i \cdot T_i + \beta_5 S_i \cdot T_i$$

This assumption says that unobservable differences in labour supply between skill/province groups can be explained by fixed group, province and time effects as well as second order interactions between them. This exploits variation (3) above, and identifies labour supply responses to changes in tax rates by comparing the responses of skill groups across provinces that were affected differently by the tax reforms. This is the assumption required for the triple-difference-in-differences estimator; the specification allows for fixed effects,

[5]For notational simplicity, I have left out a vector of observable covariates X_{it} which can easily be incorporated into this analysis

skill/province, province/time, and skill/time specific trends, but assumes that, in the absence of the tax reforms, the skill profiles of labour supply outcomes would not shift differently over time between provinces. In other words, this specification asks whether skill/province groups that receive relatively larger tax cuts increase labour supply relatively more.

Conditional on this assumption, three way interactions of skill, province and time are validly excluded from the reduced form equation and are potentially instrumental variables. The other condition for them to be valid is that after controlling for fixed effects and second order interactions, there still be variation left in MTRs. That is, we require changes it tax rates for skill groups to be significantly different between provinces or we will suffer from weak instrument issues.

In terms of the identifying assumptions, those maintained in the triple-DD model are weaker than in previous section that considered DD models. Recall that in the DD model, we require the assumption that the outcome variable be vertically parallel between provinces over time in the absence of the tax reforms. In the triple-DD model, we allow for province specific trends, potentially increasing the robustness of the estimator. However, there is a trade-off since, if the DD assumption were correct, we throw out variation (1) from above. If this is the case, then we can set $\beta_4 = 0$ and use two sources of variation to identify the response to tax reforms. The first source is the same as in the above DD section, it uses the variation in tax changes between provinces. The second source is provided by the additional variation from the differential tax changes between province/skill/time groups. Using both sources of information we are less likely to run into weak instrument problems as long as the average change in taxes are significantly different between provinces. Similarly, if we instead set $\beta_5 = 0$, then we can use variation (2) from above in addition to (3). This additional source comes from the differential changes in taxes between skill groups within provinces. This assumption is the same as in Blundell et. al. (1998), that in the absence of the tax reforms the difference in labour supply between groups can be explained by skill and time fixed effects.

4.2.1. *identifying assumptions.* The previous section made clear how exogenous variation in MTRs induced by tax reforms can be used to identify the coefficient on marginal tax rates by using groups defined by skill, province and time and imposing structure on the unobservable error term. However, further assumptions are needed in order to put the previous discussion to use. Specifically, we require group status itself not change in a non-random way in response to the tax reforms. In terms of the province groups, we assume that provincial residence is not correlated with tax reforms. This assumption does not impose that individuals do not change provincial residence or that provincial residence is not influenced by tax rates. Rather, it assumes that if an individual does relocate between provinces that this decision is not related to particular tax reforms and that if tax rates influence location then individuals are sorted before the reforms. Unfortunately, without access to longitudinal data the plausibility of this assumption cannot be empirically tested. However, considering the short period of analysis (two years prior to reforms and two years after), it does not seem out of the question.

Assumption 3. *Provincial residence does not change in a non-random way due to tax reforms.*

Skill groups have yet to be defined. As mentioned previously, due to differing progressivity of provincial tax schedules and the fact that many of the marginal tax cuts were targeted at specific income levels, skill groups should be related to income in order to exploit the variation discussed in the previous section. However, we cannot use income groups themselves since groups status would likely be endogenous and potentially correlated to tax reforms. Instead, I use observable exogenous characteristics known to be correlated with income to group individuals. Namely, individuals are grouped by age and education cells (hereafter, referred to as "skill" groups). Individuals in different cells are likely to be affected differently by tax reforms (especially since many of the reforms were concentrated at the top of the income distribution), and the same age/education cells are likely to be treated differently between provinces due to the different progressivity and social tastes in various provinces. In addition, these observable characteristics are unlikely to be influenced by tax reforms.

4.3. **implementing the estimator.**

4.3.1. *data issues.* Unfortunately, data limitations complicate things in practice. The LFS does not include information on MTRs or income information used to calculate them. For this I turn to the SLID data that has detailed information on earnings. Using a program that calculates taxes based on all of the relevant parameters used by the actual provincial and federal tax systems I calculate taxes for all individuals who fit my sample selection criteria for the LFS sample. That is, for all wage workers between ages 20-59 who are working (excluding self-employed). With these MTRs in hand, the next step is to construct skill groups.

Due to smaller sample sizes in the SLID data set, skill groups based on age/education groups proved to be impractical due to some cell sizes being small. Since these cells are important for estimating skill group MTRs, small cell sizes resulted in imprecise estimates. To overcome this difficulty while maintaining the same approach, I implement the following procedure. I run OLS regressions of age (5 year category indicators), and education (5 indicators) and their interactions along with a married indicator and provincial dummy variables on the log real wage using period 0 SLID data only. Again using the period 0 data only, I form quintiles using the predicted log wages from the distribution of all the provinces, but set the coefficients on the provincial dummies to 0 when predicting wages. Then using the coefficients from the wage regression (except provincial dummies), I predict log real wages for all individuals in the sample (including LFS data and period 1 observations) and categorise individuals based on the quintile cut-off points from the period 0 SLID predicted log real wage distribution. Thus, the number of skill groups is the number of quintiles estimated. The trade off here is that too many quantiles results in cell sizes being too small, while too few quintiles results in over grouped data that fails to capture the differential change in tax rates across the skill distribution between provinces. Below, I use deciles, which maintained reasonable sized cells.

This procedure (referred to below as the basic specification) results in skill groups that have the same observable characteristics over time and between provinces. However, since average wages, incomes and returns to observable characteristics can differ between provinces, this classification scheme risks comparing individuals that may not be in the same federal tax brackets. This may cause problems during estimation since identification hinges differential changes in taxes by province and *income* groups. Thus, I check for robustness by using alternative groupings. These include adding broad occupation classifications into the log wage regressions, increasing their predictive power, using a subset of the provinces, and using a predicted wage index instead of groups. I also use a specification that allows returns to observable characteristics to differ between provinces. This results in groups that are the same within provinces over time but whose characteristics may differ between provinces; this allows for higher correlation between federal tax treatment of groups between provinces. I will discuss these in greater detail later on.

Note that if I would have just grouped individuals based on actual log wages, the skill groups would not have been exogenous and classification could potentially change in response to tax reforms. However, using predicted log real wages based on exogenous period 0 characteristics overcomes these issues. I perform these procedures separately for males and females.

4.4. **estimation.** The above discussion suggests a simple method to estimated the reduced form equation. We can use a two-stage method where in the first stage we regress using OLS the following equation using observations from the SLID data only:

$$MTR_{it} = c + \alpha_0 P_i + \alpha_1 S_i + \alpha_2 T_i + \alpha_3 P_i \cdot S_i + \alpha_4 P_i \cdot T_i + \alpha_5 S_i \cdot T_i + \alpha_6 P_i \cdot S_i \cdot T_i + \eta_{it}.$$

I then predicted $M\hat{T}R_{it}$s for the LFS observations to use in various regressions where the dependent variables are a margins of labour supply. For example, for the hours worked per week conditional on participation we can run the following OLS regression on LFS data only.

$$H_{it} = c + \delta M\hat{T}R_{it} + \beta_0 P_i + \beta_1 S_i + \beta_2 T_i + \beta_3 P_i \cdot S_i + \beta_4 P_i \cdot T_i + \beta_5 S_i \cdot T_i + \epsilon_{it}.$$

When the margin of labour supply is participation, the second stage is replaced with a probit version. The estimated coefficient, $\hat{\delta}$, has no behavioural interpretation in terms of substitution or income effects since this reduced form model is not derived from economic theory. Rather, the interpretation is the total response of labour supply (hours of work conditional on working, for example) to MTRs.

5. Data

5.1. **LFS.** Data for this paper are drawn from two sources: The Labour Force Survey (LFS) and the Survey of Labour and Income Dynamics (SLID), both of which are public use files. The main analysis uses the LFS for the years 1999 to 2002, which brackets the tax reforms. I use additional years to address alternative hypothesis. The LFS is a monthly survey that obtains individual level labour market and demographic characteristics. The survey follows households for 6 months at a time, and excludes full-time members of the military, inmates, and those living on Indian Reserves.

I create a time series of cross sections from the LFS data. So as not to include the same individual in the sample more than once, I use only data from the months of April and October of each year. The sample includes all individuals from ages 20 to 59 who are not unpaid family workers. For the analysis on hours worked per week, I exclude self-employed individuals and focus on employees only. The sample sizes in any given year are quite large. For example, the number of observations on Alberta in 2001 is 10640. Smaller provinces are over represented compared to large ones. For this reason, in all estimates reported below, the LFS sampling weights were used.

A major drawback of the LFS is that it does not contain any information on annual income, and, therefore, tax rates are impossible to calculate using these data. I circumvent this issue by complementing the LFS with the SLID.

5.2. **SLID.** The Survey of Labour and Income Dynamics was originally designed as a longitudinal data set. However, the objectives of the Survey were extended to replace the Survey of Consumer Finances (SCF) and now releases cross-sectional versions of the data annually and contains information on families, individuals and income. The data on annual income is very detailed, with individuals having the choice of an income interview or granting permission for the survier to use T1 income tax data. The sample selection criteria I use is identical to that of the LFS.

Using both the family and personal SLID files, I construct a data set containing detailed annual income data and family structure. I then run this data through a Canadian tax calculator to create the key variable of interest: individual marginal tax rates. I then impute these values in the LFS by using variables common to both data sets.

5.3. **Descriptive Statistics.**

5.3.1. *SLID: Changes in MTRs.* This section describes the tax rates calculated from the SLID data set. Important for the following analysis is that there were changes to marginal tax rates from period 0 (1999/2000) to period 1 (2001/2002) that vary by province and income. Figure (1) plots kernel smoothed changes in tax rates for employees by province over income in year 2000 dollars. The vertical lines in the graph roughly indicate the federal tax brackets as of the year 2001. As the figure indicates, changes in tax do vary substantially by province and income level. In particular, compared to Alberta or BC, Ontario and Manitoba changes in tax rates were much less, and resulted mostly from changes to the federal tax parameters. For all three provinces, changes in rates were smaller for those in the first federal income tax bracket compared to those in higher tax brackets. At very low income levels, Albertan's had comparatively greater tax relief than BC or Ontario, but solely due to a higher exemption rate, which was $12,000 in 2001. At an income above that level, this advantage disappears, and Albertans in the first federal tax bracket received the smallest changes in taxes out of all provinces.

Past the first tax bracket, Alberta and BC changes in tax rates follow each other rather closely, although BC changes were slightly greater. These tax change profiles then diverge again past the last federal tax bracket. Ontario and BC's MTR changes track each other at lower income levels before diverging past the second tax bracket and are almost vertically parallel past the third bracket. BC's tax reform was mainly to eliminate surtaxes, thus affecting higher income earners comparatively more than lower ones, and giving the MTR change profile its downward sloping shape. While downward sloping, it is not smooth, due to the introduction of two additional tax brackets at the provincial level and the elimination of two tiers of surtaxes. By contrast, Alberta's profile is more smooth save bumps at the federal level. In all three cases, the profiles trend upwards at the last federal tax bracket. Workers in this last bracket received no federal reduction in taxes, resulting in this shape.

Figures (2) and (3) again show kernel smoothed MTR changes by province for men and women, respectively. The horizontal axis is real income percentiles. The vertical lines again roughly indicate federal tax brackets. Figure (2) shows that the majority of the men fall into the middle bracket (between \$61,509 and \$100,000 in 2001), receiving reductions in MTRs between periods of between 5% to 9%, depending on province. Figure (3) indicates that women are much more likely to fall into the first federal income tax bracket, receiving, on average, much smaller reductions in tax rates than men. Both figures again illustrate that for men and women, reductions in MTRs vary substantially by province and income level.

6. Labour Supply Responses

6.1. **Difference-in-differences.** Tables 1, 2, 3, and 4 present the results for the difference-in-differences estimates. The first two tables describe the provincial tax reforms effect on participation for women and men, respectively. The second two tables deal with hours worked conditional on participation. Each of the three columns in each table refer to different periods of comparison. For example, column (1) in table 1 compares 1999 to 2000. Each of the rows make different treatment group/control group comparisons.

It is not clear from theory which years to compare. For example, in principle one could compare any year proceeding the tax reform to any year afterwards. However, in practice, the results may be sensitive to which years are used. This could be because of a violation to assumption 1, or could simply reflect an adjustment process in labour supply behaviour to the new tax change. In the first case, one region could be hit particularly hard by a macro economic shock in a given year relative to Alberta. In this case, this differential time effect would violate the assumption and bias results. Including more time periods may reduce this effect by averaging over them. In the second case, it may just be that workers need time to adjust behaviour. For workers who are not free to set their own hours, an adjustment might mean finding a new hours/compensation bundle at another job. If this is the case, the 2001 tax reforms effects might not happen until later years.

To deal with this issue, I estimate the DD models for three different time frames that bracket the 2001 provincial tax reforms. Column (1) uses data for the year just prior to the tax reform and one year after. While column (3) uses a three year before/after bracket. All regressions include controls for age, education, age of youngest child, and marital status. The estimated coefficients can be interpreted as the total response to the policy change. For example, in the hours equation, the estimated coefficient is the average total change in hours worked per week among employed individuals in response to the tax reform, conditional on the controlled variables.

The results are fairly robust across years. First, looking at the participation results, none of the treatment effects are significantly different from zero and in most cases are the wrong sign. In fact, at the ten percent significance level, several coefficients with the wrong sign are statistically significant. Conditional on the DD assumptions, these results imply that the tax reform had little effect on participation.

Tables 3 and 4, which have the same layout, report largely the same findings. When the treatment group is Alberta, many of the treatment effects are the correct sign, but none are significant. Using BC as the treatment group, the results are largely wrong signed and some even significant. Again, it would seem that these tax reforms have little if any effect on labour supply.

However, we must use caution when interpreting results for at least two reasons. First, the assumption used for DD is a strong one, and it could be that this assumption fails. Certainly, finding the opposite results when the treatment group is Alberta compared to BC is discouraging. Secondly, even if the DD assumption were valid, the results are surely bias downward considering even the control groups received on average significant tax rate reductions due to tax rate changes at the federal level. What is more, for income earners in the first federal tax bracket in either the Ontario or Manitoba control groups received almost as large of reduction (or *larger*) in rates as the treatment groups (see figure (1)).

As a first pass at this second issue, we can redefine the treatment group to be those that are "higher income earners" living in BC or Alberta. Figure (1) indicates that the largest differences in tax rate changes occur past the second federal income tax bracket. To examine if this "treated" group had significant labour supply response, I compare them to similar individuals in Manitoba and Ontario where tax cuts were less for higher income earners. Since I do not observe incomes in the LFS, I use the SLID data to estimate a probit model where the dependent variable is equal to one if an individual has a real income level that would put them past the second income tax bracket using age, education (and their interactions), marital status, and age of youngest child as explanatory variables. Since these variables are common to both data sets, I predict the probability that an LFS observation is past the second tax bracket. I use only period 0 data for these estimates and impute the probability that period 1 individuals would be beyond this bracket had their returns to characteristics been the same as in period 0.[6]

Table (5) shows the results considering only higher predicted income earners. The table has a similar layout as the previous ones, where the control province is listed in the far left column and the different years considered listed along the top row. For Alberta, the estimated coefficients mostly have the wrong sign but are insignificant for both men and women. For BC, the results are much the same. Thus, using only higher income earners as the treatment group does not change any of the previous results; the effect of the tax reforms did not significantly affect the hours worked per week of employed individuals.

Clearly, this very coarse classification of treatment/control over simplifies the actual tax rate changes induced by the tax reforms. Tax rates changed not only by province, but by income within provinces. Grouping by province, as in these simple DD models, obviously misses out on much of the variation induced by the reforms, and is perhaps unlikely to find many effects due to miscategorisation of treatment/control groups. In addition, these DD results are conditional on strong functional form assumptions, that, if invalid, will bias the estimates. In the next section, I attempt to overcome these issues by estimating reduced form labour supply models that more fully exploit the variation induced by the provincial tax reforms.

[6]In practice, I use a real income of $30,000 year 2000 dollars as a cutoff point. I also tried $60,000 (roughly the cutoff for the third bracket). The results for men did not change significantly, while predicted samples for women where too small. I also tried stratifying by the 75th and 90th log real wage percentiles, which produced results that where not significantly different then those reported below.

6.2. **Reduced form Models.**

6.2.1. *first stage.* Before estimating labour supply models, it is necessary to first group individuals into "skill" groups using the predicted log wage procedure described in a previous section. Ontario, British Columbia, Saskatchewan, Manitoba, and Alberta are used for all analysis below. After individuals are grouped into predicted log wage decile, I proceed to the first stage estimation.

The next step is to estimate the first stage regression that fits the calculated MTRs to period/province/skill fixed effects plus all two-way and three-way interactions among them and a vector of controls that include indicators for age of youngest child, and marital status. Depending on the assumed structure of the error components in the second stage, the first stage interactions are excluded instruments. It is important that the excluded instruments be highly correlated with MTRs, thus all tables of estimates report the p-value of the test that the excluded instruments are jointly zero. In order for the excluded instruments to be valid, we must reject these hypotheses.

Figures (4) and (5) show the change in predicted MTRs (as estimated above) by skill group and province for men and women, respectively, using the basic specification for grouping. As discussed earlier, the change in MTRs between period 0 (1999/2000) and period 1 (2001/2002) favoured higher income earners. This is reflected by the downward slope in the Alberta and BC plots, but is more pronounced for males than for females. This is largely to do with the fact that at the largest predicted decile of the female log wage distribution, average incomes are still much lower compared to males. Again, this indicates that the Alberta and BC reforms benefited men relatively more than women.

In terms of the previous discussion on identification, I estimate the reduced form models based on three assumptions. The first is the Triple-DD assumption, which excludes the three way province/time/skill interactions from the labour supply models. From the figures, this source of identification comes from differences in MTR changes between provinces for each skill group - represented by the vertical distance between the provincial lines in figures (4) and (5). The null hypothesis for the test of relevant instruments in the triple-DD estimates is that there is no difference between these lines. Under the second assumption, the source of identification includes the first, but adds to it the difference in average changes in MTRs between provinces - represented by the average vertical difference between lines in Figures (4) and (5). The third assumption includes the first source of identification in addition to the change in MTRs between skill groups. In the figures, this source is represented by the downward slope of the plotted change in tax rates. This source is similar to that used in Bludell et. al. ((1998)).

6.2.2. *The parameter estimates: hours per week.* This section presents the reduced form parameter estimates. It is organised as follows: I first present estimates from the basic reduced forms separately for Men and Women. Then I present estimates using alternative methods of grouping the "skill levels" to assess robustness. All of the tables have a similar layout. The columns denoted by (1), (2), and (3) refer to estimates under alternative assumptions regarding the error term as discussed above. The first column uses the triple-DD assumption that excludes only three way skill/time/province interactions from the second stage, the second also excludes provincial/time interactions, while the third adds provincial/time interactions and excludes skill/time interactions. In these basic specifications, both the estimated coefficient and the implied elasticity are shown under each assumption. The last two rows of each panel show the p-value of the tests regarding the excluded instruments. The first of these rows is the test of relevance or the explanatory power of our instruments. Rejection of this test casts doubt on inference regarding the parameter estimates due to weak instrument problems. The second of these rows presents a heteroskedasticity-robust test of overidentifying restrictions. This tests that the instruments are validly excluded from the second stage regression. Here, failure to reject is evidence that our instruments are validly excluded.[7] All specifications include a vector of demographic characteristics in addition to the fixed provincial/time/skill effects. These include indicators for age of youngest child (4 categories) and marital status.

Table (6) shows the results using BC, Alberta, Saskatchewan, Manitoba and Ontario (referred to as all provinces) using the basic skill groups discussed above. The first panel shows the results for women. Turning first to column one, which shows the estimates from the triple-DD estimation. The coefficient on MTRs is negative but insignificant at any reasonable level. The coefficient is -0.039 with an implied elasticity of -0.033, suggesting that a one percentage increase in MTRs reduces hours worked per week by about 0.03%. This suggests that MTRs do not significantly affect hours worked per week of employed women. However, the relevance test of the instruments is highly insignificant suggesting

[7]This test is performed using a regression based approach suggested by Wooldridge (2002) (p. 123-124).

that the three way interactions of skill/time/provinces are not good predictors of MTRs. This probably renders these estimates useless, and I discuss possible reasons for this shortly.

The second and third column of the same panel report the estimates under alternative assumptions regarding the error term. In column (2), provincial/time interactions are also excluded from the second stage, and identification comes from differential changes in MTRs as it impacts each provincial/skill group as well as average differences in changes in MTRs between provinces. Under this specification, the relevance and the overidentification tests, reported in the last two rows of the panel, are satisfactory. The relevance test suggests that the instruments have much higher predictive power in the first stage, while the test of the exclusion restrictions suggests that provincial/time interactions are validly excluded from the second stage. The coefficient on MTRs under this specification is wrong signed, but insignificant; again suggesting that MTRs do not significantly affect hours worked per week. Column (3) reports much the same, however the quality of the instruments is again in question. Identification in this specification is obtained from differences in changes of MTRs between skill/province groups as well as differential changes in taxes between skill groups (time/skill).

Turning now to the bottom panel of the same table shows the results for men. The estimates are nearly the same as for women. Column (1) reports a negative coefficient, but again the significance of the instruments is in question. Columns (2) and (3) report insignificant estimates as well, and the relevance test suggest that the model is better identified. Taking all of the results from this table into account, the data suggests that MTRs have no significant effect on hours worked per week for men or women. Below I discuss the robustness of these results under alternative samples and skill groupings, but first I speculate on the possible cause of the lack of identification of the triple-DD model.

Figure (1) shows that the changes in MTRs between provinces varies by income level. In particular, the greatest differences occur at the upper end of the income distribution. The identification of the triple-DD estimator depends on how well the skill/province groups capture the vertical distance between the provincial lines in that graph. However, holding observable characteristics constant (as we do with the skill groups), incomes still vary considerably by province. Tables 7 and 8 show mean real income levels by skill group and province. The table indicates that for several skill groups, a group may be in a different federal tax bracket depending on province - thus, not capturing the vertical distance in that graph. In addition, while log wages are highly correlated with MTRs, the predictors in the log wage regressions only have an R-squared of about 0.25, which further erodes the performance of skill groups in capturing the variation on MTRs by province.[8] Below, I attempt to improve the performance of this method using a sub-sample of provinces and alternative methods to group individuals.

In table (9) I perform the same exercise, this time restricting the provinces to BC, Alberta and Ontario. Since income differences are much smaller between the predicted log wage decile groups with these provinces, it was hoped that the instruments would perform better. Unfortunately, the triple-DD restriction still does not identify the model well. However, assumptions (2) and (3) are well identified and the predicted coefficients for both men and women on MTRs are very similar to the previous table using all provinces. Again, this table suggests that there is no negative effect of MTRs on hours worked for either employed males or females.

In table (10) I again use all of the provinces, but add occupation as a grouping variable. The assumption that individuals not change groups due to the tax reform is more tenuous with this specification. However, I use ten very broad occupation categories to mitigate this problem. Adding these to the log wage regression improves the fit considerably. Column (1) shows the triple-DD estimates, which are more negative than in the previous tables but still not significant. For women, the estimated coefficient is -0.102 with an implied elasticity of -0.087, while for men the estimates are -0.123 and -0.116 respectively. However, while the power of the three way interactions is greater then in the previous specifications, their joint significance is still barely rejected at the ten percent level, indicating a weak instrument problem. However, a few time/province/skill interactions were individually significant in the first stage.

Table (11) includes occupation as a grouping variable, but restricts the sample to BC, Alberta and Ontario. The results are largely invariant to the dropping of Saskatchewan and Manitoba. In column (1), the triple-DD estimates are slightly smaller in magnitude compared to the previous table. The test of the relevance of the instruments is just barely accepted at the ten percent level for women and five percent level for men, indicating that the province/skill/time interactions performed better in the first stage. However, the weak instrument problem still persists with this specification. In columns (2) and (3) the model is again well identified, but the estimated coefficients are still small and insignificant. For

[8]I also attempted to group individuals on predicted earnings decile. This did not improve the results

men, column (3) indicates that skill/time interactions may not be validly excluded from the second stage.

The next specification check allows groups to vary in observable differences between provinces but not over time within provinces. To do this, I repeat the log wage regressions but allow the returns to observable differences to vary between provinces using period 0 data only. I predict log wages using all coefficients (including provincial dummies) and place observations into skill groups based on deciles of the entire period 0 predicted log wage distribution. While this grouping strategy provides skill groups that have more equal log wages within skill groups between provinces (and hence, more equal incomes), one might worry that observable differences between groups might confound the labour supply analysis. To mitigate such issues, I include a set of age and education indicators in the hours equation. Table (12) reports the results. Although the instruments performed slightly better, the results are virtually the same as the previous tables.

Instead of grouping, the next table shows results based on a “skill index”. For this exercise, I take the predicted log wages from the basic specification (no occupation) and standardise them to have a zero mean and unit variance (separately for men and women) and include this index in the regressions instead of groups. This places an extra restriction in the model because it forces the effect of “skill” on MTRs to be linear. The triple-DD model is then identified if the slope coefficients on province/time/skill index interactions are significant. Table 13) shows the results using this method. The first panel uses all provinces while the second uses only BC, Alberta, and Ontario. For women, this extra restriction does not improve the identification of the triple-DD estimator and the overidentifing tests for valid instruments are largely rejected. However, the results of the estimates are similar to the previous models. For men, this restriction does not alter the results significantly from previous estimates.

The last panel of this table shows one final specification check. Here, I use an alternative instrument in a procedure similar to Rothstein (2005a). Taking the skill groups based on the basic specification (and using all provinces) I create a variable called the “intended tax change” by running period 0 individuals through the tax calculator as if they were facing the period 1 tax schedule (after adjusting nominal variables for inflation). The result is the tax rate that would have occurred had individuals not re-optimised and altered behaviour. That is, it is the tax rate that each individual would have faced had they not changed their labour supply. I then collapsed the data into provincial/skill means and first differenced the resulting cells. While the change in actual tax rates by province/skill is likely to be endogenous, the "intended tax change" is thought to be exogenous since it is based only initial behaviour and the change in the tax parameters facing skill groups. Thus, I estimate the following model:

$$\Delta y_{ps} = \beta_1 \cdot P + \beta_2 \cdot S + \delta \cdot \Delta MTR_{ps} + \epsilon_{ps}$$

where P and S are province and skill group dummies, respectively, and the subscripts ps refer to province/skill cells. The change in MTRs is instrumented by the "intended tax change" which is highly correlated to the actual tax change. The estimated coefficient is shown in the last panel of table (13), which is negative but insignificant and well within the range of estimates of the previous models.

To check for heterogeneous responses, I interact the MTR variable with age of youngest child indicators. The results are shown in tables (14) and (15). Both tables show the results using skill groups with and without occupation, table (14) use the sample of all provinces while table (15) excludes Saskatchewan and Manitoba. The first stage of these estimates corresponds to those in tables (6)-(11). Standard errors are estimated with a bootstrap over both stages of estimation, using a 100 replications.

Focusing first on column (1) of table (14), which uses the triple-DD assumption, the results indicate that for both men and women, the responses to MTRs vary considerably for those with and without children. While estimates of those without children are not significant, when interacted with indicators for presence of young children they become relatively larger and significant. The negative response is larger for women than for men and decline as children age, as expected. Adding occupational indicators to the skill groups reduces the magnitude of some of the estimates for women, but does not significantly alter the results.

Columns (2) and (3) show the results with alternative exclusion restrictions. The results are very robust across assumptions, although for those with no children a few estimates are significant and wrong singed. This largely disappears when groups include occupations. Table (15) shows the results for a subset of the provinces. The results are robust, the responses of hours worked per week to MTRs are greater for those with children, larger for women than for men, and decline as children age.

6.2.3. *parameter estimates: participation.* This section presents the reduced form estimates for participation. The first stage regressions are the same as in the above section. However, the second stage is a probit model where the dependent variable is an indicator equalling 1 if an individual is observed working and 0 otherwise. The sample includes all individuals aged 20-59 who were not unpaid family workers. The layout of the tables are similar to those with the hours estimates.

Table (16) shows the results for all provinces while table (17) includes only BC, Alberta, and Ontario. The columns again refer to estimates based on the alternative identifying assumptions discussed above. While the discussion above indicates that the estimates in (1) are not well identified by the instruments, columns (2) and (3) perform much better. In any case, the results are robust across the different assumptions and suggest that MTRs do not significantly affect employment. For both men and women, coefficient estimates are small and insignificant. For men in table (17) in column (3) the estimates are wrong signed and significant at the 10% level.

This fact might indicate a problem with the model. Perhaps to due the fact that the simple reduced form model does not take into account such things a fixed costs to work or because, as Rothstein ((2005b)) suggests, what matters for participation is the average tax rate not the marginal tax rate. In some of his specifications, the coefficient on MTRs are also wrong signed in his participation models (Rothstein 2005a). Again, this section suggests that the effect of MTRs on participation are not significant.

6.3. **Summary of Results.** The above section attempted to identify responses of hours of work per week and participation to MTRs by using variation in marginal tax rates caused by provincial tax reforms. Estimates were performed under several plausible assumptions regarding instruments for MTRs and with a variety of specifications. In general, the results indicate that labour supply was not responsive to changes in MTRs in Canada during the period from 1999 to 2002. For women, estimated elasticities of hours of work per week with respect to MTRs were in the range of -0.087% to 0.054% - none of which were significantly different from zero. For men, the range was -0.104% to about 0.046% which were again not significant. Interacting MTRs with indicators for age of youngest child uncovered heterogeneous responses. For example, it was found for both men and women that those with children have larger negative responses to MTRs. These were larger for women than for men, and declined with older children. However, for those without children, the responses of men and women were virtualy idenical.

Comparing these findings with other studies is difficult, since most studies estimate behavioural models and the elasticities are directly comparable. In addition, most studies only consider married women, while the sample considered here was more general. However, a recent paper by Rothstein (2005) estimates reduced form specifications similar to the ones considered in this paper. He restricts his analysis to U.S. women only, but finds no response in hours worked per week or participation to MTRs. However, he does find significant effects of participation to the average tax rate. Moffit (1998) also considers a U.S. natural tax experiment and estimates labour supply equations of affluent men using difference-in-differences instrumental variables strategies. While his specifications are not directly comparable to those used here, he does estimate equations where the MTR is not interacted with wages. He finds no responses of hours worked per year of higher income men to reductions in MTRs, and speculates that it might be due to the fact that they already work more hours relative to other groups of men.

7. Conclusion

This paper uses variation in marginal tax rates induced by provincial tax reforms to identify the the effect of marginal tax rates on labour supply. The provincial reforms can be thought of as a "natural experiment" that provide exogenous variation in MTRs that can be used to overcome traditional problems associated with estimating labour supply elasticities. Since classical economic theory leads to mostly ambiguous prediction of these labour responses, empirical estimation of such effects is of great interest - especially in Canada where there has been few such studies.

The reforms in question were brought about by the 2001 tax-on-income changes in the federal/provincial tax structure, allowing provinces to set their own MTRs and income brackets. Thus, the methods used in this paper exploit before/after comparisons of labour responses to changes in tax rates. As a first pass at the problem, a simple difference-in-differences estimator was used before moving on to a more parameterised model that more fully exploits the variation in MTR changes between provinces over different income groups. These impacts are estimated using several different identifying assumptions.

Two main conclusions are reached. First, the results indicated that effect of MTRs on hours worked per week are quite small and not significantly different from zero. Interacting MTRs with age of youngest child uncovers heterogeneous responses. For those workers with younger children, MTR elasticities are negative and, while still moderate, are larger than for those without children or older children. This is not an unexpected result and is common in the literature.

The second finding is that men and women have very similar responses to MTRs. A traditionally held belief in the empirical literature is that labour supply elasticities for women are larger than for men. Although this position has been challenged,[9] Heckman (1993) in a survey of the literature concludes that female responses are larger. However, he goes on to say that "whether labor supply behavior by sex will converge to equality as female labor-force participation continues to increase is an open question" (p.118). The findings reported here seem to support the view that they have indeed converged for those without children, while responses for those with children still tend to be larger for women.

The implication for policy is clear. While some argue for tax reductions based on efficiency claims, I find no evidence of this. The 2001 tax reforms in B.C. and Alberta had no significant impact on participation or hours of work among employees, despite relatively large MTR reductions at upper income levels. This conclusion suggests two areas for further research. The implicit assumption made in this paper labour supply elasticities are common among income groups. However, one explanation for the lack of labour supply effect found here might be because the tax reductions were concentrated among those at the top of the income distribution - a group that may be less responsive because of already high participation rates and hours worked or because of income effects that were not examined here. Future research might relax this assumption, which may give insight into how to better design tax policy. In addition, the analysis above only considered employees, while a significant fraction of the labour force is self employed. Since the self employed arguably have more control over their hours of work, this may be an important omission.

References

Blundell, Duncan, and Meghir (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66, 827–861.

Blundell, and MaCurdy (1999): *Labor Supply: A Review of Alternative Approaches,*chap. 36, pp. 1560–1689, Handbook of Labour Economics Volume IIIA. North Holland.

[9]For example, Morz (1989) - who finds that the behaviour of married women resembled that of prime aged males

Eissa, N. (1995): "Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment," NBER Working Paper No. 5023.

Gruber, J. (1996): "Diability Insurance Benefits and Labour Supply," National Bureau of Economic Research.

Heckman (1996): "Comment on Labor supply and the economic recovery Tax Act of 1981 by Eissa, N.," *Empirical foundations of household taxation*, pp. 32–38.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence form Evaluationg a Job Training Programme," *The Review of Economic Studies*, 64, 605–654.

McKenzie (2000): "Provincial Tax Priorities in a Global and National Economy: Whats Good for the Goose Is Good for the Gander," *Canadian Tax Journal*, 48, 356–373.

McMillan, M. (2000): "Albertas Single-Rate Tax: Some Implications and Alternatives," *Canadian Tax Journal*, 48, 1019–1052.

Moffit, and Wilhelm (1998): "Taxation and Labour Supply Decisions of Affluent Men," NBER 6621.

Rothstein, J. (2005a): "The Mid-1990s EITC Expansion: Aggregate Labor Supply Effects and Economic Encidence," Princeton University.

——— (2005b): "The Mid-1990s EITC Expansion: Aggregate Labor Supply Effects and Economic Encidence," Princeton University.

Todd, P. E., and J. Smith (2003): "Does Matching Overcome Lalondes Critique of Nonexperimental Estimators?," Working Paper.

Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT press, Cambridge Massachusetts.

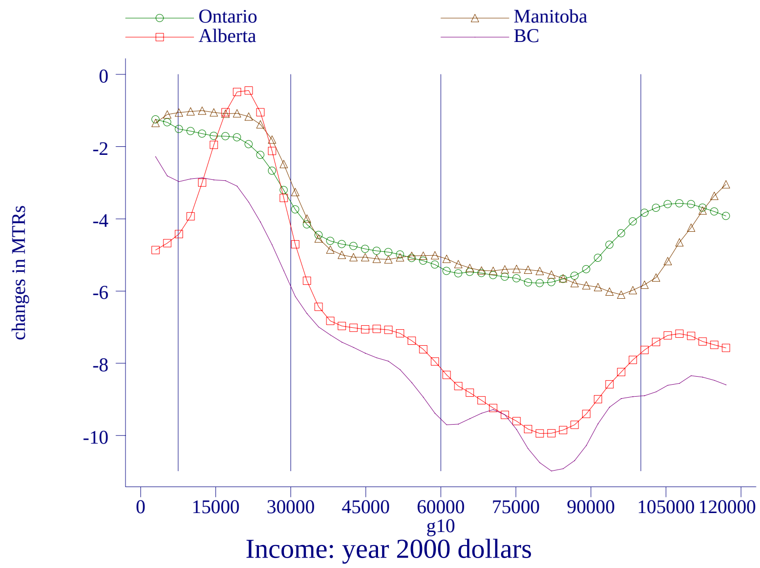

FIGURE 1. Changes in MTRs by province

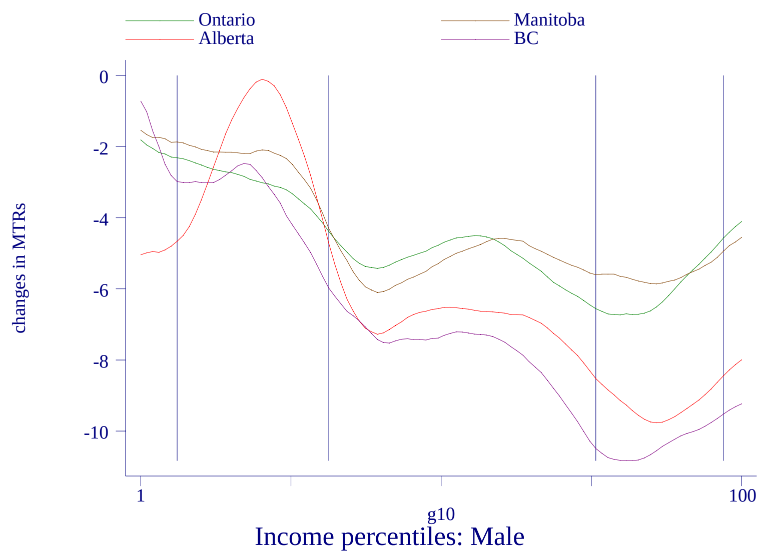

FIGURE 2. MTR changes by income percentile

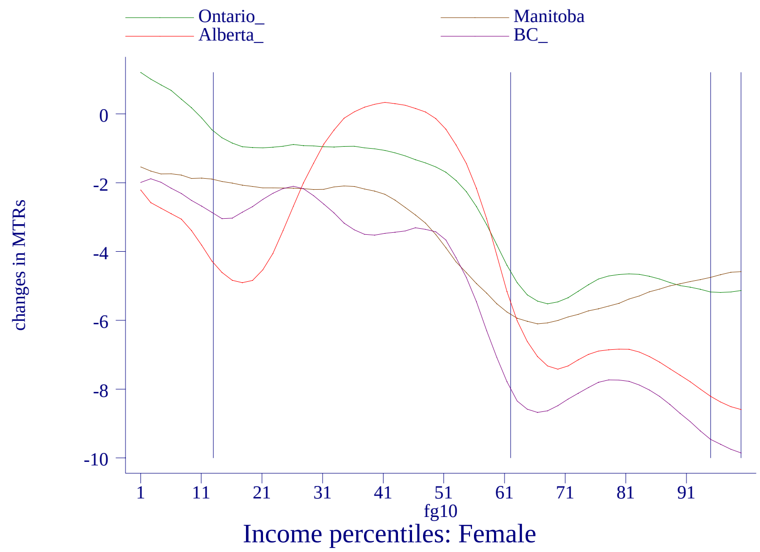

FIGURE 3. MTR changes by percentile

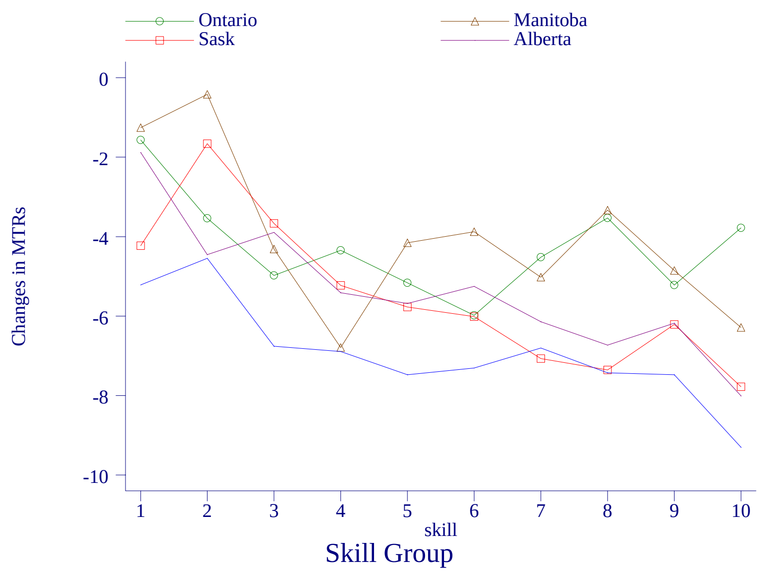

FIGURE 4. Changes in MTRs for men by province

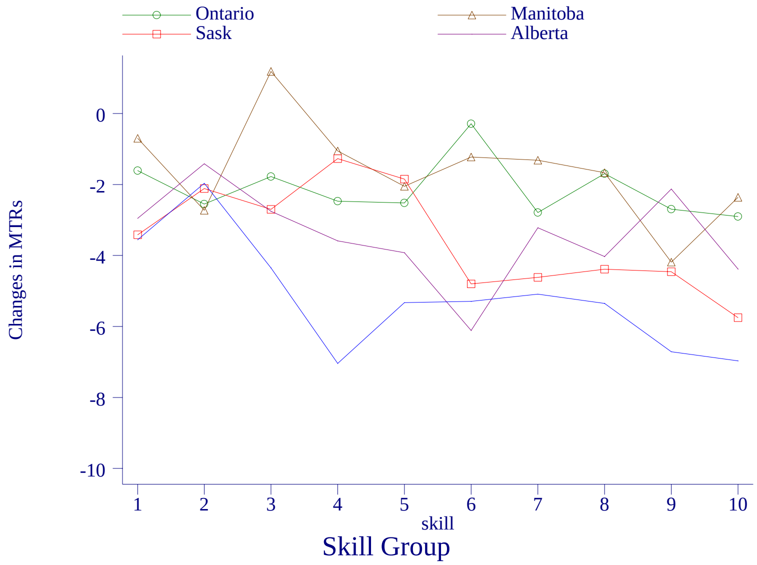

FIGURE 5. Changes in MTRs for women by province

Table 1. DD - Participation: Women

Alberta		(1) 2000/2001	(2) 1999/2002	(3) 1998/2003
vs Ontario	coef	-0.008	-0.002	-0.006
	s.e.	0.010	0.007	0.006
	p-val	0.455	0.811	0.285
vs Manitoba	coef	0.002	-0.007	-0.009
	s.e.	0.012	0.009	0.007
	p-val	0.883	0.417	0.185
vs Both	coef	-0.007	-0.002	-0.007
	s.e.	0.010	0.007	0.006
	p-val	0.490	0.757	0.258
BC				
vs Ontario	coef	-0.018	-0.010	-0.011
	s.e.	0.010	0.007	0.006
	p-val	0.062	0.160	0.066
vs Manitoba	coef	-0.009	-0.016	-0.014
	s.e.	0.013	0.009	0.008
	p-val	0.489	0.081	0.058
vs Both	coef	-0.018	-0.010	-0.011
	s.e.	0.010	0.007	0.006
	p-val	0.069	0.137	0.055
Alta/BC	coef	-0.013	-0.007	-0.009
vs Both	s.e.	0.008	0.006	0.005
	p-val	0.090	0.214	0.047

Table 2. DD-Participation:Men

		(1)	(2)	(3)
Alberta		2000/2001	1999/2002	1998/2003
vs Ontario	coef	0.011	0.011	0.007
	s.e.	0.008	0.006	0.005
	p-val	0.155	0.047	0.123
vs Manitoba	coef	0.005	-0.002	0.000
	s.e.	0.009	0.007	0.005
	p-val	0.555	0.806	0.978
vs Both	coef	0.011	0.010	0.007
	s.e.	0.008	0.005	0.005
	p-val	0.164	0.065	0.148
BC				
vs Ontario	coef	0.002	0.002	0.002
	s.e.	0.008	0.006	0.005
	p-val	0.821	0.702	0.599
vs Manitoba	coef	-0.002	-0.012	-0.006
	s.e.	0.012	0.008	0.007
	p-val	0.851	0.155	0.407
vs Both	coef	0.001	0.001	0.002
	s.e.	0.008	0.006	0.005
	p-val	0.853	0.836	0.682
Alta/BC	coef	0.005	0.005	0.004
vs Both	s.e.	0.006	0.004	0.004
	p-val	0.397	0.279	0.304

Table 3. DD-hours per week:Women

Alberta		(1) 2000/2001	(2) 1999/2002	(3) 1998/2003
vs Ontario	coef	0.289	0.110	0.188
	s.e.	0.288	0.204	0.165
	p-val	0.316	0.589	0.255
vs Manitoba	coef	0.043	-0.063	0.069
	s.e.	0.369	0.261	0.212
	p-val	0.907	0.810	0.744
vs Both	coef	0.273	0.099	0.180
	s.e.	0.284	0.201	0.163
	p-val	0.336	0.623	0.268
BC				
vs Ontario	coef	-0.320	-0.370	-0.337
	s.e.	0.285	0.200	0.165
	p-val	0.262	0.065	0.041
vs Manitoba	coef	-0.476	-0.477	-0.409
	s.e.	0.368	0.259	0.212
	p-val	0.195	0.065	0.054
vs Both	coef	-0.336	-0.381	-0.346
	s.e.	0.281	0.197	0.163
	p-val	0.232	0.053	0.034
Alta/BC	coef	-0.066	-0.171	-0.114
vs Both	s.e.	0.220	0.155	0.127
	p-val	0.763	0.269	0.367

Table 4. DD-hours per week

Alberta		(1) 2000/2001	(2) 1999/2002	(3) 1998/2003
vs Ontario	coef	-0.046	-0.105	-0.139
	s.e.	0.263	0.184	0.151
	p-val	0.861	0.567	0.359
vs Manitoba	coef	0.077	-0.013	-0.037
	s.e.	0.346	0.243	0.197
	p-val	0.825	0.958	0.852
vs Both	coef	-0.034	-0.097	-0.130
	s.e.	0.259	0.182	0.149
	p-val	0.894	0.592	0.384
BC				
vs Ontario	coef	-0.349	-0.269	-0.190
	s.e.	0.237	0.172	0.143
	p-val	0.141	0.119	0.186
vs Manitoba	coef	-0.225	-0.170	-0.079
	s.e.	0.329	0.235	0.192
	p-val	0.493	0.468	0.679
vs Both	coef	-0.345	-0.263	-0.183
	s.e.	0.233	0.170	0.141
	p-val	0.139	0.121	0.196
Alta/BC	coef	-0.211	-0.190	-0.161
vs Both	s.e.	0.190	0.136	0.112
	p-val	0.266	0.161	0.150

Table 5. DD - Only High predicted Incomes

		Men			Women		
		(1)	(2)	(3)	(1)	(2)	(3)
Alberta		2000/2001	1999/2002	1998/2003	2000/2001	1999/2002	1998/2003
vs Ontario	coef	-0.135	-0.186	-0.186	-0.237	-0.358	-0.358
	s.e.	0.272	0.191	0.191	0.529	0.414	0.414
	p-val	0.619	0.331	0.331	0.654	0.386	0.386
vs Manitoba	coef	-0.287	-0.230	-0.230	0.239	0.123	0.123
	s.e.	0.357	0.254	0.254	0.777	0.659	0.659
	p-val	0.421	0.365	0.365	0.759	0.851	0.851
vs Both	coef	-0.040	-0.162	-0.162	-0.326	-0.277	-0.277
	s.e.	0.272	0.189	0.189	0.526	0.429	0.429
	p-val	0.883	0.393	0.393	0.535	0.519	0.519
BC							
vs Ontario	coef	-0.444	-0.244	-0.244	-0.005	-0.190	-0.190
	s.e.	0.243	0.176	0.176	0.511	0.362	0.362
	p-val	0.067	0.164	0.164	0.993	0.600	0.600
vs Manitoba	coef	-0.533	-0.396	-0.396	-0.527	-0.309	-0.309
	s.e.	0.335	0.242	0.242	0.686	0.560	0.560
	p-val	0.112	0.102	0.102	0.442	0.581	0.581
vs Both	coef	-0.421	-0.222	-0.222	0.122	-0.226	-0.226
	s.e.	0.239	0.174	0.174	0.515	0.369	0.369
	p-val	0.078	0.202	0.202	0.813	0.539	0.539
Alta/BC	coef	-0.331	-0.217	-0.217	-0.169	-0.158	-0.158
vs Both	s.e.	0.193	0.139	0.139	0.416	0.306	0.306
	p-val	0.087	0.118	0.118	0.684	0.605	0.605

Table 6. Reduced Form: 5 Provinces

Women

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.039	-0.033	0.068	0.057	0.007	0.006
s.e.	0.078	0.066	0.048	0.040	0.066	0.056
p-value	0.618	0.618	0.155	0.155	0.915	0.915
relevance	0.440		0.000		0.240	
exclusion	0.648		0.557		0.577	

Men

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.037	-0.035	0.057	0.053	0.020	0.018
s.e.	0.074	0.070	0.048	0.045	0.050	0.047
p-value	0.615	0.615	0.240	0.240	0.695	0.695
relevance	0.293		0.000		0.000	
exclusion	0.712		0.723		0.430	

Table 7. Real income by Skill and Province: Women

Skill Group		Ontario	Manitoba	Sask.	Alberta	BC
1	real income	16171.01	14681.9	13226.05	14092.61	13022.85
	n	848	260	295	394	273
2	real income	21563.41	17962.47	16470.51	15615.12	19356.62
	n	823	235	164	209	221
3	real income	25469.53	20029.53	19966.29	21209.21	23729.73
	n	965	229	187	283	252
4	real income	27268.1	24817.8	22034.23	25713.12	26208.6
	n	720	162	141	204	179
5	real income	29622.9	24969.31	23986.82	25678.25	26822.27
	n	850	175	205	235	206
6	real income	28370.34	25580.02	25599.64	26587.15	26917.12
	n	853	192	223	234	251
7	real income	31387.52	27107.78	25942.38	28996.34	27265.42
	n	746	235	199	212	276
8	real income	29872.51	26499.15	27058.63	27040.46	29513.98
	n	744	216	164	227	228
9	real income	35579.21	30959.85	30464.78	31046.55	32253.94
	n	826	186	175	237	233
10	real income	46685.79	38434.7	40636.89	40839.55	42207.63
	n	837	162	135	212	181

Table 8. Real Income by skill and province:Men

Skill Group		Ontario	Manitoba	Sask.	Alberta	BC
1	real income	23167.36	21420.78	21094.89	22588.69	23522.37
	n	757	250	259	377	314
2	real income	33969.49	26378.39	24891.43	32832.23	34568.18
	n	761	212	172	246	212
3	real income	39765.1	33928.13	32009.96	38539.22	39389.34
	n	857	233	224	230	174
4	real income	40244.44	33595.29	34137.99	39687.27	41138.38
	n	916	190	189	235	221
5	real income	45538.23	38074.05	37995.78	45959.21	42134.71
	n	855	230	181	245	198
6	real income	50298.33	40444.8	39457.41	44433.18	46822.92
	n	939	210	187	245	278
7	real income	51532.71	46325	48471.94	54542.43	47708.1
	n	955	205	174	292	276
8	real income	51811.33	43715.86	45622.86	53485.19	51820.25
	n	799	167	128	215	213
9	real income	58255.66	47037.95	48769.43	51773.89	54830.18
	n	779	174	124	197	215
10	real income	71311.76	57047.02	61461.23	74007.96	67528.5
	n	782	136	159	191	208

Table 9. Reduced form: BC, Alberta and Ontario

Women

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.015	-0.013	0.064	0.054	0.016	0.013
s.e.	0.077	0.065	0.050	0.042	0.068	0.057
p-value	0.843	0.843	0.195	0.195	0.813	0.813
relevance	0.041		0.000		0.068	
exclusion	0.91		0.89		0.91	

Men

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	0.038	0.036	0.088	0.083	0.041	0.038
s.e.	0.080	0.075	0.052	0.049	0.056	0.053
p-value	0.636	0.636	0.092	0.092	0.466	0.466
relevance	0.191		0.000		0.003	
exclusion	0.87		.942		0.83	

Table 10. Reduced Form: All Prov. /w Occupation

Women

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.102	-0.087	0.024	0.020	-0.053	-0.045
s.e.	0.073	0.062	0.047	0.040	0.053	0.045
p-value	0.161	0.161	0.619	0.619	0.322	0.322
relevance	0.129		0.000		0.000	
exclusion	0.764		0.528		0.393	

Men

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.123	-0.116	0.015	0.014	0.009	0.009
s.e.	0.070	0.066	0.047	0.044	0.049	0.046
p-value	0.080	0.080	0.744	0.744	0.849	0.849
relevance	0.104		0.000		0.000	
exclusion	0.780		0.556		0.152	

Table 11. Reduced Form: BC, Alberta, Ontario /w occ.

Women

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.071	-0.060	0.045	0.038	-0.027	-0.023
s.e.	0.095	0.080	0.057	0.048	0.065	0.055
p-value	0.453	0.453	0.430	0.430	0.673	0.673
relevance	0.097		0.000		0.001	
exclusion	0.253		0.157		0.181	

Men

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.111	-0.104	0.035	0.033	0.034	0.032
s.e.	0.084	0.079	0.052	0.049	0.054	0.051
p-value	0.189	0.189	0.497	0.497	0.531	0.531
relevance	0.048		0.000		0.000	
exclusion	0.393		0.180		0.045	

Table 12. Reduced Form: Skill varies by province

Women

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.026	-0.022	0.059	0.050	0.008	0.007
s.e.	0.080	0.068	0.050	0.042	0.062	0.052
p-value	0.744	0.744	0.231	0.231	0.898	0.898
relevance	0.027		0.000		0.003	
exclusion	0.566		0.505		0.666	

Men

	(1)		(2)		(3)	
	Coef	Elast.	Coef	Elast.	Coef	Elast.
estimate	-0.059	-0.056	0.049	0.046	0.001	0.001
s.e.	0.082	0.077	0.053	0.050	0.057	0.054
p-value	0.473	0.473	0.356	0.356	0.983	0.983
relevance	0.068		0.000		0.000	
exclusion	0.845		0.696		0.948	

Table 13. Alternative Specifications

		Women (1)	Women (2)	Women (3)	Men (1)	Men (2)	Men (3)
All provinces	estimate	-0.110	0.099	0.054	-0.120	0.066	0.029
	s.e.	0.128	0.049	0.074	0.114	0.047	0.047
	p-value	0.393	0.043	0.468	0.293	0.160	0.534
	relevance	0.252	0.000	0.000	0.084	0.00	0.00
	exclusion	0.104	0.095	0.072	0.725	0.655	0.493
BC, Alberta Ontario	estimate	-0.102	0.095	0.073	-0.102	0.095	0.073
	s.e.	0.139	0.055	0.083	0.139	0.055	0.083
	p-value	0.465	0.087	0.377	0.465	0.087	0.377
	relevance	0.080	0.000	0.000	0.039	0.00	0.00
	exclusion	0.068	0.032	0.033	0.574	0.571	0.463
Intened tax change	estimate	-0.039			-0.055		
	s.e.	0.140			0.144		
	p-value	0.780			0.705		
	relevance						

Table 14. Reduce Form: Interactions

		Women			Men		
		1	2	3	1	2	3
No Children	est.	0.088	0.168	0.139	0.006	0.092	0.084
	s.e.	0.071	0.051	0.059	0.072	0.044	0.046
Child< 3	est.	-0.169	-0.169	-0.168	-0.104	-0.104	-0.102
	s.e.	0.027	0.028	0.027	0.021	0.025	0.025
Child3-5	est.	-0.310	-0.311	-0.310	-0.114	-0.115	-0.109
	s.e.	0.026	0.027	0.030	0.028	0.029	0.025
Child6-12	est.	-0.252	-0.253	-0.251	-0.056	-0.056	-0.047
	s.e.	0.021	0.022	0.021	0.023	0.022	0.023
Child13-18	est.	-0.180	-0.180	-0.180	-0.091	-0.091	-0.089
	s.e.	0.023	0.027	0.027	0.031	0.027	0.025
		With Occupation					
No Children	est.	-0.037	0.083	0.019	-0.078	0.056	0.076
	s.e.	0.065	0.047	0.048	0.067	0.048	0.045
Child< 3	est.	-0.142	-0.142	-0.141	-0.142	-0.142	-0.140
	s.e.	0.025	0.023	0.023	0.020	0.019	0.020
Child3-5	est.	-0.232	-0.232	-0.231	-0.142	-0.142	-0.138
	s.e.	0.025	0.022	0.027	0.025	0.027	0.026
Child6-12	est.	-0.146	-0.146	-0.145	-0.085	-0.085	-0.079
	s.e.	0.017	0.019	0.016	0.021	0.022	0.019
Child13-18	est.	-0.118	-0.117	-0.116	-0.065	-0.066	-0.061
	s.e.	0.023	0.022	0.024	0.026	0.025	0.023

Table 15. Reduced Form: Interaction. Alberta, BC, Ontario

		Women			Men		
		1	2	3	1	2	3
No Children	est.	0.115	0.172	0.152	0.082	0.122	0.103
	s.e.	0.078	0.057	0.067	0.087	0.055	0.051
Child<3	est.	-0.177	-0.177	-0.176	-0.125	-0.125	-0.120
	s.e.	0.029	0.025	0.027	0.025	0.027	0.024
Child3-5	est.	-0.316	-0.316	-0.315	-0.119	-0.119	-0.111
	s.e.	0.030	0.033	0.026	0.031	0.028	0.031
Child6-12	est.	-0.269	-0.270	-0.268	-0.052	-0.053	-0.039
	s.e.	0.022	0.024	0.021	0.026	0.022	0.026
Child13-18	est.	-0.190	-0.191	-0.190	-0.085	-0.085	-0.074
	s.e.	0.025	0.030	0.027	0.034	0.032	0.036
		With Occupation					
No Children	est.	-0.007	0.104	0.041	-0.055	0.082	0.108
	s.e.	0.082	0.059	0.065	0.079	0.052	0.046
Child<3	est.	-0.134	-0.135	-0.134	-0.157	-0.158	-0.156
	s.e.	0.024	0.023	0.022	0.021	0.019	0.018
Child3-5	est.	-0.245	-0.245	-0.244	-0.151	-0.152	-0.148
	s.e.	0.028	0.028	0.026	0.028	0.029	0.026
Child6-12	est.	-0.147	-0.148	-0.146	-0.089	-0.090	-0.084
	s.e.	0.022	0.022	0.018	0.021	0.023	0.020
Child13-18	est.	-0.127	-0.126	-0.124	-0.072	-0.073	-0.067
	s.e.	0.023	0.024	0.023	0.023	0.026	0.028

TABLE 16. Participation: All Provinces

Women

	(1)		(2)		(3)	
	Coef	elast	Coef	elast	Coef	elast
estimate	0.000	0.015	0.000	0.001	0.002	0.049
s.e.	0.002	0.048	0.001	0.030	0.001	0.040
p-value	0.754	0.754	0.977	0.977	0.228	0.228

Men

	(1)		(2)		(3)	
	Coef	elast	Coef	elast	Coef	elast
estimate	0.002	0.094	-0.001	-0.024	0.002	0.069
s.e	0.002	0.068	0.001	0.041	0.001	0.037
p-value	0.167	0.167	0.553	0.553	0.061	0.061

TABLE 17. Participation: BC, Alberta, Ontario

Women

	(1)		(2)		(3)	
	Coef	elast	Coef	elast	Coef	elast
estimate	-0.001	-0.020	0.000	-0.015	0.001	0.026
s.e.	0.002	0.048	0.001	0.032	0.001	0.041
p-value	0.676	0.676	0.639	0.639	0.527	0.527

Men

	(1)		(2)		(3)	
	Coef	elast	Coef	elast	Coef	elast
estimate	0.002	0.080	-0.001	-0.036	0.002	0.078
s.e.	0.002	0.075	0.001	0.044	0.001	0.042
p-value	0.282	0.283	0.423	0.423	0.061	0.061