Abstract

Do progressive marginal income tax rates discourage self-employment? We assume risk neutrality to construct an implicit surtax on stochastic income relative to steady income, arising from a convex tax schedule. It is computed as part of a structural probit model with earnings equations and a tax simulator. The tax convexity variable and the net-of-tax income difference between self- and paid-employment have the predicted signs and high levels of statistical significance for the probability of self-employment. A simulated flat tax reform suggests the tax effects are small.

JEL: H24, J24, C35

An Empirical Model of Tax Convexity and Self-Employment*

Jean-François Wen and Daniel V. Gordon

March 4, 2013

*Wen: University of Calgary; Gordon: University of Calgary and University of Stavanger. We thank Gustavo Bobonis, Herb Schuetze, Michael Smart, and seminar participants at the Department of Finance Canada for helpful comments on an earlier draft. We are indebted to two anonymous referees whose comments substantially improved this paper. Wen gratefully acknowledges support from the Social Sciences & Hunamities Research Council grant no. 410-2005-0526.

I Introduction

Entrepreneurship is regarded as vital for generating employment and innovation. It is also a risky occupation with high potential returns and high failure rates. Since the income tax system alters the relationship between risk and reward, its design may be an important determinant of entrepreurial activity. For example, a revenue compensated increase in the rate of a linear progressive tax is expected to raise the share of the workforce engaged in entrepreneurship.[1] A plausible view, however, is that progressivity achieved with a rising *marginal* tax rate schedule is likely to discourage entrepreneurship because the higher tax rates penalize success by more than the lower tax rates provide relief against poor returns.

Self-employment is frequently used as a proxy for entrepreneurship in empirical studies. The proxy is imperfect since only a portion of self-employed individuals ever innovate or hire employees. The policy implications of this fact are addressed later. Nevertheless, one way to examine the tax penalty hypothesis empirically is to estimate the effects of the tax system on the choice between self-employment and employed labor. In this paper, we develop a structural model of earnings and discrete occupational choice with a focus on our "tax convexity" variable characterizing the size of the tax penalty on activities that result in fluctuating incomes under marginal tax rate progression.

Tax convexity is measured by Gentry and Hubbard (2000, 2005) as the *spread* in the marginal (or average) tax rates for successful and unsuccessful self-employment outcomes, defined relative to the ability of each individual. It is clear why high tax rates in the event of success may discourage entrepreneurial risk-taking. It is less evident why low tax rates in the case of failure would discourage it. Gentry and Hubbard provide two illustrations of why the size of the downside component of tax convexity may be negatively correlated with self-employment rates. First, an imperfect loss offset in the event of failure corresponds to a low tax rate applied to losses. Second, since the alternative to self-employment is to work for wages, a comparatively low tax rate in the bracket encompassing wage earnings increases the opportunity cost of self-employment. In both of these examples, small tax rates in the middle to lower tax brackets discourage self-employment. Although the examples show the potential relevance of Gentry and Hubbard's tax convexity variable as a negative correlate of self-employment rates, the approach depends on particular configurations of self-employment and wage employment earnings. Loss offsetting, for example, is not relevant to an entrepreneur earning low but positive profits.

A more direct way to measure tax convexity is to calculate the expected value of the tax liability of an entrepreneur facing a distribution of possible returns and to compare this burden with the same individual's tax liability at their predicted income. If high marginal tax rates on successful entrepreneurs penalize success more than low marginal tax rates provide a form of indemnity in the case of business failure, then the expected tax liability will exceed the tax liability of the predicted income. The implicit surtax on risky income may discourage self-employment relative to paid employment, which usually provides a more stable source of income.[2] This idea is implemented in our empirical work. Specifically, we calibrate person-specific income distributions from an estimated self-employment earnings equation and we use these in conjunction with a tax simulator to compute a tax convexity variable for each person. The variable expresses the implicit surtax as a proportion of net income (consumption). The predicted negative impact of tax convexity on self-employment assumes risk-neutrality.[3] Another important concept in our model is the net-of-tax income difference between self-employment and employment for every individual evaluated at their predicted incomes from the earnings equations.

The effects of progressive taxation on the probability of self-employment are estimated with a probit model using Statistics Canada's cross-sectional public-use microdata files from the Survey of Labour and Income Dynamics (SLID) over the years 1999-2005. The period spans an important income tax reform in Canada in 2001. The Canadian Tax and Credit Simulator (CTaCS) developed by Milligan (2007) is used to calculate the total federal and provincial tax and transfer implications of the employment/self-employment choice for each individual in our data. The tax parameters in CTaCS include federal and provincial income taxes, payroll taxes, tax credits, and income transfers and their clawbacks, all of which matter for tax convexity and net-of-tax differentials.

Previous work on the effects of tax convexity on occupational choice has focused exclusively on the United States. The income tax system in Canada differs in important respects from the U.S. tax system. The personal and corporate income tax systems are integrated in Canada for incorporated businesses earning up to $500,000 (in 2012). This is achieved with a personal dividend tax credit that offsets corporate taxes.[4] In contrast, in the U.S. the tax structure provides an incentive to shift labor income to corporate income when income realizations are large.[5] There exists a $500,000 lifetime capital gains exemption on small business shares in Canada, which has no equivalent in the U.S. Marginal tax rates are higher in Canada and the rates apply to individual taxpayers, rather than to married couples as in the U.S. These and other comparative features of the Canadian tax system provide an alternative fiscal setting to explore the effects of tax progressivity on self-employment.

Studies on the effects of taxes on self-employment have mainly estimated reduced form equations, which specify the self-employment choice as a function of variables thought to determine labor market outcomes, such as education, experience, age, capital, occupation, marital status, parenthood, and labor market conditions, along with marginal or average tax rates.[6] A less common approach utilizes structural models derived formally from the self-employment decision as a function of the earnings differential between self-employment and wage work. Only the studies by Bruce (2000), Parker (2003), and Fossen (2007, 2008) estimate structural models containing earnings equations and *net-of-tax* income differentials in place of gross income differentials as determinants of self-employment; none include measures of tax convexity. As shown by Gentry and Hubbard (2000, 2005), the omission of tax convexity considerations may be important for understanding how taxation affects occupational choice.[7]

We find that the probability of self-employment is negatively related to our tax convexity variable and is positively related to the net-of-tax income differential between self-employment and employment, both with high levels of statistical significance.[8] However, the sizes of the impacts of the tax variables appear to be small. This observation is based on the outcome of a policy simulation with our estimated model. We simulated the effect of changing the Canadian federal income tax in year 2000 to a flat tax rate of 20 percent. The projected increase in the average self-employment rate from the tax reform is about 3 percent for individuals in the sample with predicted annual incomes above $50,000 and 1 percent for the working population as a whole.

The rest of the paper is organized as follows. Section 2 explains the theory that guides our empirical model. Section 3 describes the data. Section 4 provides the empirical specifications and discusses issues of self-selection and identification. Section 5 presents the results. Section 6 reports the hypothetical flat tax simulation. Section 7 concludes.

II Theory

A Progressive Tax Function

Denote individual i's before-tax income in occupation j by y_j^i and after-tax income by x_j^i. Represent income tax progressivity using the function (Musgrave and Thin, 1948, and Benabou, 2000):

$$x_j^i = \left(y_j^i\right)^{1-\tau} \widehat{y}^{\tau} \tag{1}$$

where $0 < \tau < 1$ and $\widehat{y} > 0$ are parameters of the tax system. At the income level $\widehat{y}$ the taxpayer's tax liability is zero. The term $1 - \tau$ is the elasticity of after-tax income with respect to before-tax income. Higher values of τ correspond to more redistributive tax systems (Jakobsson, 1976). The expression "tax convexity" conveys the fact that the tax liability function $T(y_j^i) \equiv y_j^i - x_j^i = y_j^i \left(1 - (\widehat{y}/y_j^i)^{\tau}\right)$ is strictly convex. Marginal tax rates are characterized by $dT/dy_j^i = 1 - (1-\tau)(\widehat{y}/y_j^i)^{\tau}$ and $d^2T/(dy_j^i)^2 = \tau(1-\tau)(\widehat{y}/y_j^i)^{\tau}/y_j^i > 0$. The implication of tax convexity is that a stochastic income generates a larger expected tax liability than the tax liability associated with the expected income.[9]

B Occupational Choice

Utility is assumed to depend on after-tax income (consumption) and on an hedonic index Q_j^i of non-pecuniary job characteristics associated with an occupation j. The index is composed as follows:

$$Q_j^i = \sum_n \delta_n^i q_{nj} \tag{2}$$

where q_{nj} is the measure of a non-pecuniary job characteristic in occupation j and δ_n^i is its utility weight. Following Rees and Shah (1986), let the utility function be

$$u_j^i = U(x_j^i, Q_j^i) = \left(x_j^i\right)^{b^i} \exp\left(d^i Q_j^i\right) \tag{3}$$

where d^i and b^i are preference parameters; risk aversion is decreasing with b^i and non-pecuniary considerations increase with d^i.

Before-tax income is assumed to be lognormally distributed:

$$y_j^i \sim LN(\mu_j^i, \sigma_j^i).$$

Substituting (1) for x_j^i into (3) gives

$$u_j^i = \left(y_j^i\right)^{(1-\tau)b^i} \widehat{y}^{\tau b^i} \exp\left(d^i Q_j^i\right). \tag{4}$$

An individual i chooses the occupation $j \in \{S, E\}$ (representing self-employment and employment, respectively) that maximizes expected utility $E(u_j^i)$. Using properties of the lognormal distribution (Aitchison and Brown, 1966, Theorem 2.1),

$$E\left[\left(y_j^i\right)^{(1-\tau)b^i}\right] = \exp\left\{(1-\tau)b^i\mu_j^i + \frac{1}{2}\left((1-\tau)b^i\sigma_j^i\right)^2\right\}.$$

Thus maximizing expected utility in (4) is equivalent to maximizing the following utility index:

$$V_j^i(Q_j^i, \overline{y}_j^i, c_j^i) = (1-\tau)\ln \overline{y}_j^i - \frac{1}{2}(1-\tau)(1-b^i(1-\tau))\ln\left(1+(c_j^i)^2\right) + \frac{d^i Q_j^i}{b^i}, \tag{5}$$

where $\overline{y}_j^i \equiv E(y_j^i) = \exp\left(\mu_j^i + \frac{1}{2}(\sigma_j^i)^2\right)$ is the mean of y_j^i and $c_j^i = \sqrt{\exp(\sigma_j^i)^2 - 1}$ is the coefficient of variation of y_j^i.

Expected utility is higher with self-employment compared with employment if $C^{*i} \equiv V_S^i(Q_S^i, \overline{y}_S^i, c_S^i) - V_E^i(Q_E^i, \overline{y}_E^i, c_E^i) \geq 0$. Using (5), $C^{*i} \geq 0$ is equivalent to

$$(1-\tau)\left[\ln\left(\frac{\overline{y}_S^i}{\overline{y}_E^i}\right) - \frac{1}{2}(1-b^i(1-\tau))\ln\left(\frac{1+(c_S^i)^2}{(1+(c_E^i)^2}\right)\right] + \frac{d^i}{b^i}\left[Q_S^i - Q_E^i\right] \geq 0. \tag{6}$$

The appearance of the term $(1-\tau)$ as a multiplicative factor on the left-hand side of (6) indicates that greater tax progressivity (higher τ) reduces the relative importance of financial differences, both reward and risk, in determining occupational choice. We can decompose the tax effects in (6) into two parts.

The first term $(1-\tau)\ln\left(\overline{y}_S^i/\overline{y}_E^i\right)$ in (6) can be interpreted as the percentage change in after-tax income that results from switching occupations from employment to self-employment, evaluated at the expected income from each occupation. In the econometric model this variable is called "NetIncdif." Hence, we would expect the estimated coefficient on NetIncdif to be positive in a probit equation for self-employment. The second appearance of a tax term is less familiar and relates to tax convexity.

C Tax Convexity

Consider the second tax term in (6):

$$-\frac{1}{2}(1-\tau)(1-b^i(1-\tau))\ln\left(\frac{1+(c_S^i)^2}{(1+(c_E^i)^2}\right). \tag{7}$$

We would normally expect $c_S^i > c_E^i$, as self-employment is associated with entrepreneurial risk-taking.[10] Set $c_E^i = 0$ to represent the relative safety of paid-employment and suppose that individuals are risk-neutral, so $b^i = 1$. In that case, the term (7) (in absolute value) reduces to an expression that has a simple interpretation. It is the increase in tax liability taken on by self-employed individuals due to the volatility of their earnings, expressed as a proportion of their disposable income:

$$\frac{E[T(y_S^i)] - T(\overline{y}_S^i)}{\overline{x}_S^i} \approx (1/2)(1-\tau)\tau \ln\left(1 + \left(c_S^i\right)^2\right), \tag{8}$$

where $T(\overline{y}_S^i)$ is the tax burden at the expected self-employment income and $\overline{x}_S^i = \overline{y}_S^i(\widehat{y}/\overline{y}_S^i)^\tau$ is the net-of-tax income at $\overline{y}_S^i$.[11] We construct the tax convexity variable in our empirical model using a logarithmic formulation of (8),

$$\text{Tax Convexity} = \frac{E[T(y_S^i)] - T(\overline{y}_S^i)}{T(\overline{y}_S^i)} \times \frac{T(\overline{y}_S^i)}{\overline{x}_S^i} \approx \ln\left(\frac{E[T(y_S^i)]}{T(\overline{y}_S^i)}\right) \times \frac{T(\overline{y}_S^i)}{\overline{x}_S^i} \tag{9}$$

and we refer to it as "Convexity." The predicted sign on Convexity is unambiguously negative (see (7)).[12]

In the case of risk aversion ($0 < b^i < 1$), the effect of tax convexity is given by (7), which contains an interaction of τ and b^i. Smaller values of b^i (greater risk aversion) weaken the magnitude of the negative impact of tax progressivity on self-employment due to the insurance effect of redistributive taxation.[13]

We assume risk neutrality in our empirical work. If agents are in fact risk-averse, then the coefficient of the Convexity variable will be biased downward in our probit regression. To see this, note that a generalized convexity measure based on (7) can be written as the sum of two terms, i.e.,

$$\begin{aligned} & (1/2)(1-\tau)(1-b^i(1-\tau)) \ln\left(1 + (c_S^i)^2\right) \\ = \quad & (1/2)(1-\tau)\tau \ln(1 + \left(c_S^i\right)^2) + (1/2)(1-b^i)(1-\tau)^2 \ln(1 + \left(c_S^i\right)^2). \end{aligned} \tag{10}$$

The first term in (10) is the same as the tax convexity measure derived under risk neutrality and given by (8). The second term is zero if agents are risk-neutral but positive if agents are risk-averse. The two terms in (10) should be positively correlated if all agents are similarly risk-averse ($b^i \approx b < 1, \forall i$) with variation in the terms stemming from c_S^i. Since the predicted sign of the coefficient on Convexity is negative, there would be negative omitted variable bias (Wooldridge, 2009: 91). This implies that if agents are risk-averse, our estimated probit regression would overstate the negative impact of the Convexity variable on the probability of choosing self-employment. However, the potential size of the bias is dampened if the coefficient of relative risk aversion, $1 - b^i$, in the second term of (10) is negatively related to the size of risk, $\ln(1 + (c_S^i)^2)$, due to self-selection into entrepreneurial activities by less risk-averse persons (see Kihlstrom and Laffont, 1979). The evidence provided by Kan and Tsai (2006) is that risk tolerant individuals are relatively more likely to become self-employed.

We do not observe the non-pecuniary aspects embodied in the hedonic indexes and d^i. Observable variables believed to affect occupational choice are used as proxies for the unobservable characteristics in the empirical model.

Finally, note that self-employment and paid-employment determine an individual's labor earnings, but the total tax liability and consumption depend also on investment income, alimony payments, and so on. In other words, total income is

$$y_j^i = e_j^i + a^i, \tag{11}$$

where e_j^i is labor earnings of i in occupation j, and a^i is i's non-labor income. We take a^i as exogenous when calculating an individual's tax liability but we predict e_j^i with econometric equations.

III Data

The model is estimated with Statistics Canada's public-use microdata files from the Survey of Labour and Income Dynamics (SLID). This is an anonymized collection of income, labour, and family variables on persons in Canada derived from the longitudinal SLID. In a given year, the longitudinal SLID consists of two overlapping panels with a new panel introduced every three years. The sample frame for the survey covers all individuals in Canada. We use annual cross-sections of individuals who report their main activity as working, covering the seven year period 1999-2005 for a total of 142,278 observations of which 8,674 are defined as self-employed. The SLID data prior to 1999 omits several explanatory variables, while 2005 was the last available year for CTaCS during our study.

The key variable of interest in our analysis is employment status. We define persons to be self-employed if they report their major source of earnings to be from self-employed work and they also report no farm income. Persons are defined to be employed if they report their major source of earnings to be from wages and salary and they also report no farm income. Thus, we construct a binary dependent variable equal to one if a person is self-employed and equal to zero when a person is employed as defined above.[14]

The SLID Person File is used to download all variables associated with the individual and this is then linked to the SLID Economic Family File to collect household information. The variables obtained from SLID and used in the empirical work are defined in Table 1. The main earnings variable measures salaries, wages and self-employment income. Two variables that characterize the financial status of individuals and are assumed to impact the self-employment choice are investment income and taxable capital gains.[15] We specify a number of control variables describing individual characteristics believed to affect labor market outcomes such as age, gender, education, disability, marital status, and number of children in the household. To account for labor market conditions we control for unemployment spells and type of employment (white collar, blue collar, and service).

[INSERT TABLE 1 HERE]

Table 2 shows summary statistics for the combined seven years. The data show an interesting statistical picture of the self-employed versus employed. On average the self-employed have lower earnings and greater variation in earnings compared to the employed. This is consistent with the general notion that self-employed work is more risky. As well, the self-employed show both higher average levels of investment income and capital gains, are slightly less educated, moderately older and more likely to be married and have more children.

[INSERT TABLE 2 HERE]

IV Empirical Model

A Constructing the Tax Variables

The variables NetIncdif and Convexity are constructed from estimated earnings equations. In general, the earnings equations can be written as,

$$\ln\left(earnings_j\right) = Z\gamma_j + \varepsilon_j \tag{12}$$

for $j =$ employed (E) and self-employed (S). Z is a vector of explanatory variables common to both equations and ε_j is an error term assumed to be normally distributed. A problem with least squares estimation of equation (12) is the possibility of omitted variable bias caused by self-selection into employed/self-employed classifications. We follow the standard Heckman (1979) procedure to correct for omitted variable bias by estimating, for each year, a reduced form probit specified over the choice between employment and self-employment.[16] From these equations the inverse Mills ratios are calculated for the self-employed and employed individuals and included in the appropriate earnings equation.[17] The corrected earnings function is estimated, separately in each year, first using data only for employed individuals, which is used to generate predicted employment earnings for each individual in the data set, both self-employed and employed. This operation is repeated by estimating the earnings equation using only the self-employment data and then predicting self-employed earnings for each individual in the data set. What we obtain from this exercise is the expected earnings for an individual if that person had chosen self-employment or alternatively paid-employment.[18] Next, the predicted earnings for self-employment and paid-employment for each individual, together with their financial and personal characteristics, are used in the CTaCS simulator to generate tax liabilities from each occupational choice. Finally, the predicted income (i.e. predicted earnings plus the non-labor income reported in the data) and tax liabilities in each occupation are used to calculate the percentage difference in after-tax income between self-employment and employment (NetIncdif). This calculation is undertaken for every individual in each year of our sample. Note that since NetIncdif is (the log of) a ratio of incomes between self-employment and employment, price inflation across years poses no problem for pooling the seven years of data to estimate the structural probit equation. NetIncdiff has a mean value of -0.249 and a standard error of 0.287. The negative average value of NetIncdiff indicates that self-employment is typically less remunerative than employment. This underscores the importance of non-pecuniary job charactersitics in the occupational choice.

Numerical simulations are used to compute Convexity. For every individual, we take 1,000 draws from a normal distribution with the conditional mean equal to the predicted log of self-employment earnings from (12) and a conditional variance of $\widehat{\sigma}_S^2\widehat{h}_S(Z)$, where $\widehat{\sigma}_S$ is the standard error from regression (12) and $\widehat{h}_S(Z)$ are the predictions of a heteroskedasticity function that is estimated using standard procedures to fit the squared non-robust residuals of (12) (Wooldridge, 2009: 283-284). The regression coefficients of the heteroskedasticity function are highly statistically significant. The assumption behind our method is that an individual's perceived variation of his or her possible earnings from self-employment is adequately explained by observable factors such as age, education, marital status, location, type of job, etc. The value of each draw is exponentiated and added to non-labor income to obtain total income. Next, we use CTaCS to calculate the corresponding tax liability in each iteration. The expected tax liability of a given individual in self-employment is a simple average from the 1,000 iterations.[19] From the same set of simulations we obtain the tax liability at the expected income $[T(\bar{y}_S^i)]$ and after-tax income $[\bar{x} = \bar{y}_S^i - T(\bar{y}_S^i)]$ to calculate Convexity in accordance with equation (9). Like NetIncdif, Convexity is a ratio, which eliminates the issue of inflation across the sample years in the pooled probit regression. Convexity has mean value of 0.011 and a standard error of 0.16. This means that the annual tax penalty due to fluctuations in self-employment income is on average 1.1 percent of disposable income. Convexity is typically largest for individuals with incomes in the $80,000 to $95,000 range, where the average value is just over 2 percent.

Variations in NetIncdif and Convexity stem from several factors besides the predictions of the earnings equations. Non-labour incomes can be large and vary substantially across individuals. Eligibility for certain tax credits depends on the taxpayer's family structure. There are also notable differences in provincial income tax schedules; e.g. Alberta has a single tax rate of 10 percent while other provinces have different rates of marginal progression. An important federal tax reform in 2001 lowered the marginal income tax rates and there have been changes to payroll tax rates for pensions and employment insurance during the sample period. Figure 1 displays the coefficient of variation (CV) of Convexity across income bins with widths of $5,000. The CV is generally between 1 and 2 across a wide range of middle incomes and exhibits spikes near the federal marginal income tax thresholds of approximately $30,000 and $60,000, as individual tax burdens around these incomes become particularly sensitive to the sizes of the earnings shocks. Overall, Figure 1 suggests that the variation within income groups is important for identifying the effect of Convexity in the structural probit model.

[INSERT FIGURE 1 HERE]

The cross-sectional data excludes information on previous occupational choices and other factors that might be relevant for self-employment such as IQ. However, as we account for self-selection bias in the earnings equations, our premise is that omitted variables are uncorrelated with NetIncdif and Convexity, which are constructed from the estimated earnings equations. The concern of potential omitted variable bias for the coefficient estimate of Convexity is addressed further with Table 3, which shows that the correlations between Convexity and other key explanatory variables in the structural probit regression are low (Wooldridge, 2009: 93).

[INSERT TABLE 3 HERE]

B Occupational Choice

Using these generated variables the structural probit equation for the probability of self-employment is written as,

$$C = Z\gamma + X\beta + \alpha_I \cdot NetIncdif \ + \alpha_C \cdot Convexity + \varepsilon_C. \tag{13}$$

The vector Z is as defined earlier. The vector X includes investment income, taxable capital gains, and the number of children in the household. ε_C is assumed to be *iid* and asymptotically normally distributed. Additional dummy variables are added to equation (13) to account for provincial differences and year effects. The probit model presumes that individuals evaluate their choice between employment and self-employment each year.

V Empirical Results

This section presents the results obtained from estimating the annual reduced form probit, the earnings functions for each year, and the structural probit for the pooled seven years of data and for the individual years.

A Reduced Form Probit

The results of the maximum likelihood estimation for the reduced form probit for 1999 to 2005 are reported in Table 4. Recall that the purpose of the reduced form probit is to obtain the inverse Mills ratios that allow correction for the omitted variable problem in the earnings equations. Consequently, to justify the reduced form equations it is necessary that they be econometrically identified and the estimation procedure must produce consistent estimates. For identification we maintain the assumption of normality of the population error terms. We default to the large number of observations used in estimation and the central limit theorem to support this assumption. However, normality is not enough for identification without at least one identifying variable in the reduced form estimation relative to the earnings equation. The reduced form equations have three identifying variables: investment income (inva27), capital gains (capgn27) and kids (number of children in the household). These variables proxy for risk in the self-employed decision but are assumed not to have a determining impact on earnings. The investment variable is statistically important at less than the 10% level in six of the seven years, the capital gains variable is statistically unimportant in all years but 1999, and the number of kids is significant at the 5% level in five of the seven years.[20]

[INSERT TABLE 4 HERE]

For parameter consistency we argue that all right-hand-side variables in the reduced form probits can be treated as exogenous variables in estimation. As our regressors represent individual characteristics, the financial outcomes of past choices, and geographical location, this assumption appears to be reasonable. Therefore, we conclude that in each year the probit equation is well specified, identified and econometrically reasonable, and we proceed to calculate the inverse Mills ratios for each earnings equation in each year.

B Earnings Equations

The maximum likelihood robust estimates for the self-employed earnings equation are reported in Table 5 and for the employed earnings equation in Table 6.[21] The base reference individual is a white collar, single female, without disability, and resident in Ontario. Comparing Tables 5 and 6 we observe a better fit to the data for the employed earnings equation compared to the self-employed earnings equation. In every year the estimated employed earnings equation shows all coefficients with p-values less than 1%. In contrast, the self-employed earnings equations report p-values greater than 10% for 14 out of 77 coefficients. This likely reflects the risk involved in this type of activity, the heterogeneity of the self-employed and the number of observations used in estimating the different earnings equations; there are approximately 1,300 observations per year for the self-employed compared to 19,000 per year for the employed. Nevertheless, over 80% of the coefficients in the self-employed earnings equations are statistically accurate at less than the 10% level and a null hypothesis that all coefficients except the intercept term are zero is easily rejected. In support of the estimates it is also worth noting that except for one occasion there are no sign changes in coefficients across the different years for both self-employed and employed equations.

[INSERT TABLE 5 AND TABLE 6 HERE]

The estimated earnings coefficients for the employed are in line with the literature. Earnings increase with years of schooling and with the male and married indicators, and (although not reported) Ontario offers the highest earnings levels relative to the other provinces. Blue collar and service workers show negative earnings relative to white collar workers. The results for the self-employed are similar except for the married indicator. While marriage is strongly positively correlated with earnings in the employment equation in all years, it is negatively correlated with self-employment earnings and is statistically significant at 5% in six of the seven years.

It is also interesting to compare the inverse Mills ratio for each equation. The results show a very strong positive bias for those individuals who have chosen either self-employment or employed labor.[22] Importantly, a Wald test of independence between the reduced form probit and earnings is soundly rejected in each year for each pair of equations. These results indicate that self-selection is important in occupational choice and that correcting for omitted variables is necessary for consistent estimation of the earnings equations.

C Structural Probit

The structural probit equation is estimated using the seven years of pooled data. The results are reported in Table 7. Over 140,000 observations are used to estimate the structural equation using the pooled time-series and cross-section data. The reference person is a white collar, single female, without disability, in Ontario in 1999. For both consistency and efficiency in estimating the parameter values, a maximum likelihood estimator is augmented by bootstrapping the standard errors with one thousand replications. The bootstrap is necessary because of the generated nature of the NetIncdif and Convexity variables.

[INSERT TABLE 7 HERE]

We observe a strong positive relationship of the probability of self-employment with investment income, age, the number of children in the household, and the indicator variables for male, married, and blue collar and service occupations. In contrast, the number of weeks unemployed and age-squared are negatively correlated with the probability of self-employment, while capital gains, years of schooling and the disability indicator have no impact.

The dummy variable for British Columbia shows a statistically positive coefficient relative to Ontario. In fact, summary statistics show British Columbia with the highest rate of self-employment at 7.04% compared to Ontario with 5.45%. The 2001 year dummy clearly picks up the substantial decline in the probability of self-employment possibly due to stock market problems in that year (e.g. reductions in day-traders and internet ventures).

The most interesting results are for the after-tax income differential and tax convexity variables reported near the bottom of Table 7. The coefficient for NetIncdif is positive with a p-value of 0.000 and the coefficient for Convexity is negative with a p-value of 0.019. The results suggest that tax progression influences the occupational choice between self-employment and paid-employment.[23] Although the coefficients of NetIncdif and Convexity have the predicted signs, their magnitudes differ. This points to possible measurement errors in the constructed variables (e.g. due to risk aversion) or to behaviors that depart from a strict adherence to the theory.[24]

As previously mentioned, higher income tax rates could increase the marginal benefit of pursuing self-employment if this is a means to under-report income. Relatedly, since many self-employed are exempt from paying portions of payroll taxes, high payroll tax rates may induce individuals to select self-employment.[25] Hence, we included two additional variables that vary by year and by province, as controls in the structural probit regression: the effective average income tax rate reported by Sharpe and Arsenault (2009: Table 10A) and the effective payroll tax rate from Wen and Wan (2011: Table A.5). The tax evasion/avoidance hypotheses imply positive coefficients on both tax rates. The effective tax rates were found to be statistically insignificant and had negligible impacts on the other coefficients.[26]

As a robustness check the structural probit equation is taken to the yearly data. Table 8 reports the coefficient results only for the net income difference and tax convexity variables. The NetIncdif and Convexity coefficients maintain their expected signs in every year. NetIncdif shows low p-values in most years, while Convexity is significant at conventional levels in two years, but exhibits high p-values in 2002 and 2003.[27] The yearly results suggest overall that the full structural probit model is econometrically reasonable.

[INSERT TABLE 8 HERE]

VI Policy Application

A way to guage the sizes of the estimated tax impacts on self-employment is to simulate a hypothetical tax reform. We examine a reform in year 2000 that would replace the graduated federal income tax rates with a single rate of 20 percent, while keeping tax credits, payroll taxes, provincial income tax rates, and all other features of the tax system unchanged. The 20 percent flat rate is chosen because it was considered in a research paper of the federal Department of Finance at that time. According to Pigeon (2001), simulations with the Department of Finance's SPSD/M model showed that a 20 percent flat rate, with no other changes to the income tax, would generate a sustainable balanced budget for the federal government.[28] Table 9 describes the actual federal marginal personal income tax schedule in 2000. The 20 percent flat rate would lower the top and middle marginal income tax rates, while raising the bottom rate.

[INSERT TABLE 9 HERE]

Table 10 shows the results of our simulation. The first row of figures gives the predicted average probability of self-employment in year 2000 under the actual tax structure; the second row is the prediction for the flat tax. The figures in each column are for a given subsample of the data—by occupation or income group—except the one labelled "Average" which is for the entire sample. The third and fourth rows of figures provide the *partial* effects of NetIncdif and Convexity, respectively, of the flat tax while holding the other variable at its predicted value with the actual tax code in 2000. For the sample as a whole, the flat tax would generate slightly less than a 1 percent increase in the average probability of self-employment. The predicted increase is 1.4 percent for white collar workers but negligible for service and blue collar workers. The impact is somewhat higher for individuals with predicted self-employment incomes greater than $50,000, where the increase is close to 3 percent.

[INSERT TABLE 10 HERE]

The partial effects allow us to better understand the reasons for the muted impacts of the flat tax. The fourth row shows the impact of Convexity under the flat tax, holding NetIncdif at the level associated with the 2000 tax code. Here we see, for example, that the reduction in tax convexity under the flat tax accounts for most of the increase in self-employment of individuals with incomes above $50,000. Furthermore, for the population as a whole, Convexity under the flat tax increases self-employment but NetIncdif under the flat tax, shown in the third row, reduces self-employment. As expected the flat tax substantially lowers the average size of Convexity (from 1.2 percent in 2000 to 0.8 percent under the flat tax, for a 33 percent reduction). The weak or adverse effect of the flat tax for NetIncdif arises because the flattening of the marginal tax schedule tends to favor employment over self-employment, since self-employment income is comparatively lower than employment income in the data. On average the magnitude of NetIncdiff increases by almost 4 percent (from -22.5 percent under the 2000 code to -23.3 percent under the flat tax). These are interesting observations that militate against the view that average self-employment rates would increase much under a flat tax. Furthermore, the finding that even the partial effect of Convexity on the probability of self-employment is quite small suggests that the disincentive for self-employment under existing progressive marginal tax rates, although present, is mild. This might be due to the social insurance effect of progressive taxation that would tend to benefit the self-employed if they are in fact risk-averse.

VII Conclusions

This paper makes two basic contributions. First, we develop a theoretical random utility model of occupational choice between self-employment and employment in the presence of progressive marginal income taxation, which guides the construction of the relevant tax variables. Our tax convexity variable is new and it has a clear interpretation as an implict surtax on risky income expressed as a proportion of disposable income. The predicted negative impact of the convexity variable on self-employment assumes risk-neutrality. Second, we estimate a structural probit model that uses generated variables defining earnings and tax liabilities for each individual, in each employment classification. An estimated heteroskedasticity function is fitted to the residuals of the self-employment earnings equation and is used to calibrate person-specific earnings distributions from self-employment based on the observed charactersitics of each individual in the sample. The tax convexity variable is constructed from numerical simulations of the self-employment earnings distributions. The model is estimated on Canadian cross-sectional data from 1999-2005.

Our major finding is that marginal tax progressivity matters statistically in the self-employment choice. The tax convexity variable is negatively related to the probability of self-employment. The difference between after-tax incomes from self- and paid-employment evaluated at the predicted incomes is also highly significant with the predicted positive sign, in contrast to most previous findings in the literature. We use the estimated model to simulate a hypothetical 20 percent federal flat tax reform in year 2000. The actual federal tax rates in 2000 ranged from 17 to 29 percent. The simulation suggests that the impacts of the tax variables on the magnitudes of self-employment are quite small. This finding corroborates the general conclusions of Bruce and Mohsin (2006) based on their time series analysis of U.S. data, but contrasts with the large effects of tax convexity estimated by Gentry and Hubbard (2000, 2005). Further analysis suggests the reasons for the muted impacts of the flat tax. Although the partial effects of tax convexity and the net income differential are both modest, a further reason why the flat tax reform fails to gain much traction toward self-employment is because the impacts of the two tax variables tend to be offsetting. The flat tax clearly reduces tax convexity, but since wage earners on average have larger incomes than the self-employed, cutting the marginal tax rate on middle incomes tends to favor paid-employment over self-employment. The flat tax reform has a more pronounced positive effect on self-employment among individuals in the sample with relatively large predicted incomes. Hence, if an objective of policy is to enhance innovation and job creation, then tax reforms may be better aimed at targeting the low frequency, but potentially highly innovative, upper income earners, via cuts in the top marginal tax rate, rather than attempting to raise the overall rate of self-employment. The impediment to self-employment caused by tax convexity could be addressed by permitting income averaging across years, a policy that was repealed in Canada in 1988 and in the United States in 1986 due to its administrative complexity.

Whether promoting self-employment is an appropriate policy goal is another matter. Many self-employed activities are unlikely to create spillovers (e.g. taxi drivers, hairdressers, small shop owners, painters, lawyers, etc.). Even in the context of innovation and growth, it is theoretically ambiguous whether there is too little or too much entrepreneurial research activity (Aghion and Howitt, 1992). Of course, a flat rate tax cut could enhance entrepreneurship in other ways besides occupational choice, such as increased investment in capital, or more hours spent working.

References

[1] Aghion, Philippe, and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica* 60:2 (1992), 323–351.

[2] Aitchison, John, and James A.C. Brown, *The Lognormal Distribution with Special Reference to its Use in Economics* (Cambridge: Cambridge University Press, 1966).

[3] Amemiya, Takeshi, "Tobit Models: a Survey," *Journal of Econometrics* 24: 1-2 (1984), 3–61.

[4] Benabou, Roland, "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review* 90:1 (2000), 96–129.

[5] Bernhardt, Irwin, "Comparative Advatntages in Self-Employment and Paid Work," *Canadian Journal of Economics* 27, 273–289.

[6] Blanchflower, David G., and Andrew J. Oswald, "What makes a young entrepreneur?" *Journal of labor Economics* 16:1 (1998), 26–60.

[7] Blau, David M., "A Time-Series Analysis of Self-Employment in the United States," *Journal of Political Economy* 95:3 (1987), 445–467.

[8] Bruce, Donald, "Effects of the United States Tax System on Transition into Self-Employment," *Labour Economics* 7:5 (2000), 545–574.

[9] Bruce, Donald, and Mohammed Mohsin, "Tax Policy and Entrepreneurship: New Time Series Evidence," *Small Business Economics* 26:5 (2006), 409–425.

[10] Cameron, A. Colin, and Pravin K. Trivedi, *Microeconometrics Using Stata*, (Stata Press, 2008).

[11] Clemens, Christiane, and Maik Heinemann, "On the Effects of Redistribution on Growth and Entrepreneurial Risk-Taking," *Journal of Economics* 88:2 (2006), 131–158.

[12] Cullen, Julie B., and Roger H. Gordon, "Taxes and Entrepreneurial Risk-Taking: Theory and Evidence for the US," *Journal of Public Economics* 91 (2007), 1479–1505.

[13] Domar, E. and R. Musgrave, "Effects of Proportional Taxes on Risk-Taking," *Quarterly Journal of Economics* 58 (1944), 388–422.

[14] Evans, D.S., and Leighton, L.S., "Some Empirical Aspects of Entrepreneurship," *American Economic Review* 79:3 (1989), 519–535.

[15] Ferede, Ergete, "Tax Progressivity and Self-Employment: Evidence from Canadian Province," *Small Business Economics* (2011), DOI.

[16] Fölster, S., "Do Lower Taxes Stimulate Self-Employment," *Small Business Economics* 19 (2002), 135–145.

[17] Fossen, F.M., "Risky Earnings, Taxation and Entrepreneurial Choice - A Microeconometric Model for Germany," German Institute for Economic Research. DIW Discussion Paper 705 (2007).

[18] Fossen, F.M., "Would a Flat Tax Stimulate Entrepreneurship in Germany? A Behavioral Microsimulation Analysis Allowing for Risk," German Institute for Economic Research. DIW Discussion Paper 773 (2008).

[19] Fossen, Frank M., and Viktor Steiner, "Income Taxes and Entrepreneurial Choice: Empirical Evidence from Two German Natural Experiments," *Empirical Economics* 36:3 (2009), 387–513.

[20] García-Peñalosa, Cecilia, and Jean–François Wen, "Redistribution and Entrepreneurship with Schumpeterian Growth," *Journal of Economic Growth* 13:1 (2008), 57–80.

[21] Gentry, William M., and R. Glenn Hubbard, "Tax Policy and Entrepreneurial Entry," *American Economic Review Papers and Proceedings* 90:2 (2000), 283 –287.

[22] Gentry, William M., and R.Glenn Hubbard, "'Success Taxes', Entrepreneurial Entry, and Innovation," in Adam B. Jaffe & Josh Lerner & Scott Stern. "Innovation Policy and the Economy, Volume 5," NBER Books, National Bureau of Economic Research, (2005).

[23] Hansson, Åsa, "Tax policy and Entrepreneurship: Empirical Evience from Sweden," *Small Business Economics 38:4* (2012), 495–513.

[24] Heckman, James J., "Sample Selection Bias as Specification Error," *Econometrica* 47:1 (1979), 153–161.

[25] Kan, Kamhon and Weis-Der Tsai, "Entrepreneurship and Risk Aversion," *Small Business Economics* 26(5): 465–474.

[26] Kanbur, S.M., "Risk Taking and Taxation: An Alternative Perspective," *Journal of Public Economics* 15:2 (1981), 163–84.

[27] Le, Anh.T., "Empirical Studies of Self-Employment," *Journal of Economic Surveys* 13:4 (1999), 381–416.

[28] Lee, Lung-Fei, "Unionism and Wage Rates in a Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review* 19 (1978), 415–33.

[29] Long, James E., "The Income Tax and Self-Employment," *National Tax Journal* 35:1 (1982), 31–42.

[30] Hurst, Erik and Lusardi, Annamaria, "Liquidity Constraints, Household Wealth, and Entrepreneurship," *Journal of Political Economy* 112:2 (2004), 319–347.

[31] Jakobsson, Ulf, "On the Measurement of the Degree of Progression," *Journal of Public Economics* 5 (1976): 161–168.

[32] Kihlstrom, Richard E. and Jean-Jacques Laffont, "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy* 87:4 (1979), 719–748.

[33] Milligan, Kevin, *Canadian Tax and Credit Simulator*, Database, software, and documentation, Version 2008–1 (2007).

[34] Moore, C.S. and Mueller R.E., "The Transition from Paid to Self-Employment in Canada: the Importance of Push Factors," *Applied Economics* 34 (2002), 791–801.

[35] Musgrave, R. and Tun Thin, "Income Tax Progression 1929-48," *Journal of Political Economy* 56:6 (1948), 498–514.

[36] Parker, Simon C., "A Time Series Model of Self-Employment under Uncertainty," *Economica* 63 (1996), 459–75.

[37] Parker, Simon C., "Does Tax Evasion Affect Occupational Choice?" Oxford Bulletin of Economics and Statistics 65:3 (2003), 379–394.

[38] Pigeon, Marc-André, "A Closer Look at the Flat Tax," Economics Division, Department of Finance, PRB 01-10E (23 August 2001).

[39] Rees, Hedley and Shah, Anup, "An Empirical Analysis of Self-Employment in the UK," *Journal of Applied Econometrics* 1 (1986), 95–108.

[40] Robson, Martin T., and Colin Wren, "Marginal and Average Tax Rates and the Incentive for Self-Employment," *Southern Economic Journal* 65 (1999), 757–773.

[41] Schuetze, Herb J., "Taxes, Economic Conditions and Recent Trends in Male Self-Employment: a Canada-US Comparison," *Labor Economics* 7:5 (2000), 507–544.

[42] Schuetze, Herb J., and Donald Bruce, "Tax Policy and Entrepreneurship," *Swedish Economic Policy Review* 11 (2004), 233–265.

[43] Sharpe, Andrew, and Jean-François Arsenault, "Living Standards Domain of the Canadian Index of Wellbeing," CSLS Research Report 2009-4 (June 2009).

[44] Stabile, Mark, "Payroll Taxes and the Decision to be Self-Employed," *International Tax and Public Finance* 11:1 (2004), 31–53.

[45] Statistics Canada, "Survey of Labour and Income Dynamics (Public-Use Microdata File)," 1999-2005.

[46] Wen, Jean-François, and Rui Wan, "Payroll Taxes in Canada from 1997-2007," Department of Economics, University of Calgary Working Paper No. 2011-14.

[47] Wooldridge, Jeffrey M., *Introductory Econometrics*, 4th edition, (Mason: South-Western Cengage Learning, 2009).

Notes

[1]An income tax can encourage risk-taking because it reduces the variance of net-of-tax income (Domar and Musgrave, 1944). Linear progressivity positively impacts entrepreneurship, innovation, and growth in García-Peñalosa and Wen (2008). In Kanbur (1981) and Clemens and Heinemann (2006) the effect of redistributive taxation on entrepreneurship is ambiguous due to the general equilibrium effects on firm sizes and wages.

[2]For example, a taxpayer in Ontario in 2012 without dependents who earns $60,000 of taxable income in one year and $20,000 the next would pay 16 percent more in income taxes compared with the same individual earning $40,000 in both years.

[3]If agents are risk-averse, they derive an insurance benefit from progressive taxation, which is entangled with the implicit surtax on risky income, as we show later.

[4]The integration arises only if the corporation pays dividends.

[5]See Cullen and Gordon (2007) for the impact of this tax feature on self-employment.

[6]See the surveys by Le (1999) and Schuetze and Bruce (2004).

[7]Our focus is the rate of self-employment, as in, e.g., Rees and Shah (1986) and Bernhardt (1994), rather than on occupational transitions, as the public-use SLID files do not facilitate longitudinal analysis.

[8]Fossen (2007), Fossen and Steiner (2009), and Hansson (2012) find negative relationships between income tax rates and self-employment. Ferede (2011) finds that greater marginal tax progressivity reduces self-employment rates in aggregate data for Canadian provinces. However, most studies conclude the contrary result, which is often attributed to the incentive for tax avoidance or tax evasion at high marginal tax rates (e.g., Long, 1982, Blau, 1987, Parker, 1996, Robson and Wren (1999), Bruce, 2000, and Schuetze, 2000).

[9]Note that the curvature of the tax schedule is increasing in τ until $\tau = 1/2$ and then it decreases. The empirically relevant range is $\tau < 1/2$ (Ferede, 2011).

[10]Gentry and Hubbard (2005) provide evidence that self-employed individuals face greater income fluctuations across years than do employed individuals.

[11]It is straightforward to show that $\left[E(T(y_S^i) - T(\bar{y}_S^i)\right] / \bar{x}_S^i = 1 - e^z$, where $z = -(1/2)\tau(1-\tau)\ln(1+(c_S^i)^2)$. Since $1 - e^z = -\left(z + \frac{z^2}{2!} + \frac{z^3}{3!} + ...\right)$, then for small z, $1 - e^z \approx -z = (1/2)\tau(1-\tau)\ln(1+(c_S^i)^2)$.

[12]We estimate coefficients of NetIncdiff and Convexity without restricting their magnitudes to be the same.

[13]Mathematically, $d^2[-(1/2)(1-\tau)(1-b^i(1-\tau))]/(d\tau db) = -(1-\tau) < 0$. The fact that tax convexity discourages self-employment *regardless* of the degree of risk aversion in (7) is due to assuming constant relative risk aversion.

[14]The vast majority of individuals in our sample report either only self-employment income or only employment income. Few individuals obtain similar amounts of income from both occupations. This supports our assumption that individuals make a binary choice between self-employment and employment. Annual national rates of self-employment in our sample are 6.30 (1999), 5.72 (2000), 5.29 (2001), 6.65 (2002), 6.48 (2003), 6.02 (2004), 6.46 (2005).

[15]Blanchflower and Oswald (1998) find evidence that financial constraints are an important deterrent to self-employment. However, the issue is controversial. Hurst and Lusardi (2004) contest the previous evidence regarding liquidity effects and show that the capital requirements for many entrepreneurial activities are small.

[16]The Heckman correction will be identified with exclusion restrictions rather than functional form.

[17]Under the assumption of normality of the error structure the inverse Mills variable is the ratio of the normal density (φ) to the cumulative normal density (Φ). For the

self-employed the inverse Mills ratio is defined as $-\varphi(\Psi)/\Phi(\Psi)$ and for employed labor it is defined as $\varphi(\Psi)/\left(1-\Phi(\Psi)\right)$ (Lee, 1978).

[18]We correct for the downward bias caused by taking the exponential of predicted log earnings (see Wooldridge, 2009: 211).

[19]See Cameron and Trivedi (2008: 133-135) for a discussion of the method. The probit estimation results are stable with 1,000 draws used to construct Convexity.

[20]Rees and Shah (1986) make a similar argument for identification using a 'child' variable for identification. However, this variable is statistically insignificant in their estimation and brings into question the identification of their reduced form equation.

[21]The error structure in the earnings equations are corrected for possible heteroscedasticity but as described by Amemiya (1984) this does not account for possible heteroscedasticity introduced by the fact that the inverse Mills ratios are generated variables. However, this problem does not cause inconsistency in the parameter estimates and consistency is what is required here.

[22]Note the inverse Mills ratio for the self-employed is negative; see footnote 17.

[23]Finding that the after-tax income differential is an important determinant of the self-employment/employment choice stands in contrast to the statistical results of Parker (2003), who concludes from examining U.K. data, that the choice of self-employment is unrelated to pecuniary considerations.

[24]A t-test of the restriction $\alpha_I + \alpha_C = 0$ in Eq. (13) is rejected with a p-value of 0.027 in the pooled data. The annual probits fail to reject the hypothesis at the 10% level except in 2001.

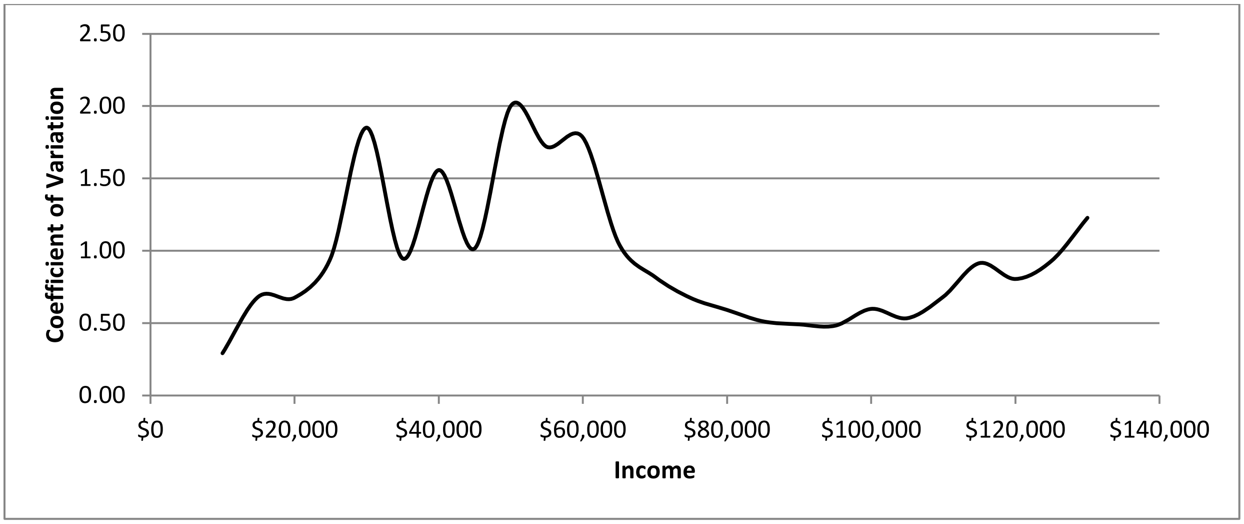
[25]Stabile (2004) and Folster (2002) find that self-employment increased after changes in legislation, exempting self-employment from certain payroll taxes.

[26]The coefficients are 0.022 for the payroll tax rate with a p-value of 0.540, and -0.010 for the income tax rate with a p-value of 0.683. The p-value for the joint significance of the two tax rates is 0.785. Hence, the variables were dropped.

[27]These two years were peak years of recession in Canada following the collapse of the tech bubble. It is plausible that the role of taxation in occupational choice is diminished when unemployment is prevalent, due to "push" factors (see Moore and Mueller, 2002). If this is the case, the impact of Convexity may be more pronounced in normal economic conditions than our reported impacts from the pooled regression.

[28]Furthermore, in Jauary 2000, delegates of the founding convention of the Canadian Alliance (a predecessor to the Conservative Party that currently forms the government in Canada) voted to make a 17 percent flat rate the cornerstone the new party's election platform. It should be noted that a single rate income tax without other changes in the tax system is a less drastic reform than a Hall-Rabushka flat tax.

Figure 1: Coefficient of Variation for Convexity in Income Bins of $5,000



Note: The Convexity variable is constructed by the authors from the self-employment earnings equations and the CTaCs program. The graph shows the coefficient of variation of Convexity for each subsample of the population with predicted self-employment incomes inside $5,000 intervals (e.g. $10,001 to $15,000, $15,001 to $20,000, etc.).

Table 1: Variable Definitions

Variable	Definition or SLID name
Self-employed	Takes value of 1 if major source of income from self-employed earnings and no farm earnings. Takes value of 0 if major source of income from wages and salaries and no farm earnings.
Earnings	Salaries, wages and self-employed earnings (earng42)
Investment income	inva27
Capital gains	capgn27
Schooling	Years of schooling (yrschl18)
Age	ecage26
Disability	Takes value 1 if disabled, 0 otherwise
Male	Takes value 1 if male, 0 otherwise
Married	Takes value 1 if married, 0 otherwise
Kids	Number of children in household
Weeks unemployed	wksuem28
Blue collar	Takes value 1 if blue collar worker, 0 otherwise
White collar	White collar worker (omitted category of worker)
Service	Takes value 1 if service worker, 0 otherwise
D00 to D05	Year dummies for 2000 to 2005
Newfoundland to British Columbia	Provincial dummies (Ontario is omitted)

Table 2: Summary Statistics: 1999-2005

	Self-Employed		Employed	
	Mean	Std. Dev.	Mean	Std. Dev.
Earnings	31,536.69	36,925.31	37,950.23	25,171.33
Investment income	961.83	5,888.37	665.56	4,559.58
Capital gains	329.63	4,001.68	263.56	3,992.94
Schooling	13.51	3.35	13.66	2.98
Age	44.51	10.09	40.33	11.24
Disability	0.15	0.36	0.14	0.35
Male	0.62	0.49	0.52	0.49
Married	0.81	0.39	0.69	0.46
Kids	0.83	1.09	0.76	1.04
Weeks unemployed	0.38	2.91	1.13	4.83
White collar	0.46	0.49	0.51	0.49
Blue collar	0.31	0.46	0.27	0.44
Service	0.23	0.42	0.22	0.41
Observations	8,677		133,652	

Note: Authors' calculations based on Statistics Canada's annual cross-sectional public-use microdata files from the Survey of Labour and Income Dynamics (SLID). The figures for White collar, Blue collar, and Service show the proportion of workers in each job category.

Table 3: Correlations between Convexity and Selected Variables, 1999-2005

Investment income	Capital gains	Schooling	Age	Disability	Male
-0.037	-0.014	0.026	0.011	0.121	0.021
Married	**Kids**	**Weeks unemployed**	**Blue collar**	**Service**	**NetIncdif**
0.038	0.014	-0.034	-0.031	0.027	-0.005

Note: The table shows the correlation between the Convexity variable, which is constructed by the authors, and other variables in the model using annual cross-sections of the Survey of Labour and Income Dynamics (SLID).

Table 4: Maximum Likelihood Estimates of

the Reduced Form Probit Equation, by Year

	1999	2000	2001	2002	2003	2004	2005
Investment	0.005	0.004	0.004	0.008	0.005	0.002	0.003
Income	(0.003)	(0.067)	(0.037)	(0.000)	(0.000)	(0.413)	(0.099)
Capital gains	0.003	0.001	0.001	0.000	0.002	0.002	0.000
	(0.056)	(0.208)	(0.801)	(0.946)	(0.319)	(0.115)	(0.936)
Schooling	-0.100	-0.098	-0.056	-0.127	-0.113	-0.078	-0.108
	(0.000)	(0.641)	(0.033)	(0.000)	(0.000)	(0.002)	(0.000)
Schooling	0.004	0.004	0.002	0.004	0.005	0.287	0.004
squared	(0.000)	(0.000)	(0.046)	(0.000)	(0.000)	(0.003)	(0000)
Age	0.512	0.523	0.614	0.408	0.427	0.286	0.183
	(0.000)	(0.000)	(0.000)	(0.005)	(0.000)	(0.142)	(0.042)
Age squared	-0.043	-0.042	-0.054	-0.033	-0.032	-0.018	-0.005
	(0.000)	(0.001)	(0.000)	(0.003)	(0.001)	(0.109)	(0.637)
Weeks	-0.033	-0.014	-0.017	-0.030	-0.012	-0.024	-0.019
unemployed	(0.000)	(0.001)	(0.003)	(0.000)	(0.001)	(0.000)	(0.001)
Disability	-0.015	-0.017	0.022	-0.010	-0.050	-0.041	0.033
	(0.722)	(0.718)	(0.625)	(0.779)	(0.208)	(0.300)	(0.340)
Male	0.166	0.220	0.222	0.201	0.118	0.123	0.162
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Married	0.152	0.129	0.134	0.154	0.106	0.125	0.113
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Kids	0.048	0.029	0.013	0.032	0.038	0.038	0.017
	(0.000)	(0.042)	(0.397)	(0.023)	(0.009)	(0.014)	(0.242)
Blue collar	0.075	0.032	0.077	0.067	0.116	0.103	0.123
	(0.068)	(0.404)	(0.057)	(0.067)	(0.000)	(0.008)	(0.001)
Service	0.209	0.212	0.265	0.222	0.201	0.117	0.119
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.002)
Constant	-2.492	-3.23	-3.398	-3.015	-3.039	-2.596	-2.549
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations							
Self-Employment	1,389	1,185	1,049	1,374	1,239	1,122	1,319
Employment	20,664	19,357	18,810	19,331	18,828	17,582	19,080

Note: The table shows the reduced form probit equations for each year, which are used to obtain the inverse Mills ratios. The dependent variable is binary with a value of 1 if the individual is self-employed and 0 if employed. P-values are shown in parentheses with robust standard errors. Provincial dummy variables are included in each equation but not reported. The data are annual cross-sections from the Survey of Labour and Income Dynamics (SLID).

Table 5: Maximum Likelihood Estimates of the

Heckman Earnings Equation for Self-Employed Individuals, by Year

	1999	2000	2001	2002	2003	2004	2005
Schooling	0.068	0.079	0.075	0.063	0.065	0.051	0.046
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	-0.245	-0.691	-0.642	-0.908	-0.657	-0.235	-0.025
	(0.298)	(0.003)	(0.015)	(0.000)	(0.000)	(0.271)	(0.900)
Age squared	0.011	0.052	0.053	0.081	0.057	0.011	-0.014
	(0.692)	(0.056)	(0.080)	(0.002)	(0.007)	(0.659)	(0.537)
Weeks	0.039	0.025	0.029	0.041	0.009	0.037	0.017
unemployed	(0.001)	(0.006)	(0.012)	(0.001)	(0.259)	(0.001)	(0.102)
Disability	-0.259	-0.179	-0.189	-0.160	-0.095	-0.125	-0.238
	(0.008)	(0.094)	(0.077)	(0.056)	(0.202)	(0.210)	(0.003)
Male	0.245	0.049	0.123	0.021	0.119	0.199	0.158
	(0.001)	(0.536)	(0.163)	(0.786)	(0.064)	(0.012)	(0.029)
Married	-0.292	-0.251	-0.127	-0.158	-0.146	-0.187	-0.227
	(0.001)	(0.008)	(0.156)	(0.057)	(0.030)	(0.029)	(0.003)
Blue collar	-0.398	-0.343	-0.392	-0.444	-0.519	-0.399	-0.396
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Service	-0.815	-0.898	-0.918	-0.942	-0.986	-0.751	-0.612
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	6.781	8.099	7.387	8.406	7.362	6.698	6.536

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mills ratio	-2.192	-2.584	-3.331	-2.022	-2.664	-3.507	-1.875
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)
Rho test	401.05	520.35	252.76	372.72	404.18	242.05	288.46
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Null test	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	1,388	1,185	1,048	1,374	1,238	1,122	1,319
Iterations	8	7	12	5	9	7	5

Note: The table shows Maximum Likelihood estimates of the earnings equation for self-employed individuals with a Heckman correction for self-selection. The dependent variable is the logarithm of earnings. The data are annual cross-sections from the Survey of Labour and Income Dynamics (SLID). Provincial dummy variables are included in each equation but not reported. P-values are shown in parentheses with robust standard errors. The Mills ratio refers to the inverse Mills ratio calculated from the reduced form probit equations. The Rho test refers to a Wald test of independence between the reduced form probit and the earnings equation. The Null test shows the p-value on a Wald test that all coefficients except the constant are equal to zero. Iterations refer to the number of iterations until convergence.

Table 6: Maximum Likelihood Estimates of

Heckman Earnings Equation for Employed Individuals, by Year

	1999	2000	2001	2002	2003	2004	2005
Schooling	0.044	0.047	0.049	0.047	0.049	0.051	0.053
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.982	1.009	0.968	0.948	1.006	1.025	0.931
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age squared	-0.103	-0.106	-0.102	-0.099	-0.105	-0.108	-0.097
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weeks	-0.031	-0.028	-0.034	-0.036	-0.030	-0.035	-0.032
unemployed	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Disability	-0.073	-0.068	-0.086	-0.056	-0.079	-0.094	-0.075
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.422	0.427	0.414	0.427	0.413	0.404	0.399
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.068	0.067	0.074	0.0542	0.052	0.057	0.048
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Blue collar	-0.130	-0.125	-0.119	-0.149	-0.159	-0.110	-0.128
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Service	-0.449	-0.466	-0.455	-0.458	-0.474	-0.452	-0.469
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.570	0.498	0.592	0.701	0.576	0.523	0.725

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mills ratio	1.410	1.642	1.813	1.274	1.333	1.349	2.237
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Rho test	42.75	44.25	30.39	33.34	41.19	32.80	42.89
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Null test	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	20,661	19,347	18,799	19,318	18,824	17,577	19,078
Iterations	6	9	6	6	6	6	7

Note: The table shows Maximum Likelihood estimates of the earnings equation for employed individuals with a Heckman correction for self-selection. The dependent variable is the logarithm of earnings. The data are annual cross-sections from the Survey of Labour and Income Dynamics (SLID). Provincial dummy variables are included in each equation but not reported. P-values are shown in parentheses with robust standard errors. The Mills ratio refers to the inverse Mills ratio calculated from the reduced form probit equations. The Rho test refers to a Wald test of independence between the reduced form probit and earnings equation. The Null test shows the p-value on a Wald test that all coefficients except the constant are equal to zero. Iterations refer to the number of iterations until convergence.

Table 7: Structural Probit Estimation of the Choice between Self-Employment and Employment using Pooled Data: 1999-2005

Variable	Coefficient	P-value
Investment income	3.04E-06	0.003
Capital gains	4.67E-07	0.698
Schooling	0.000	0.973
Age	0.053	0.000
Age squared	-0.043	0.000
Disability	0.003	0.825
Male	0.187	0.000
Married	0.139	0.000
Kids	0.024	0.000
Weeks unemployed	-0.026	0.000
Blue collar	0.097	0.000
Service	0.205	0.000
Newfoundland	-0.195	0.000
Prince Edward Island	-0.014	0.691
Nova Scotia	-0.096	0.000
New Brunswick	-0.089	0.000
Quebec	-0.065	0.000
Manitoba	-0.043	0.058
Saskatchewan	-0.023	0.337

Alberta	-0.052	0.016
British Columbia	0.094	0.000
D00	-0.043	0.031
D01	-0.091	0.000
D02	0.004	0.854
D03	-0.008	0.667
D04	-0.011	0.681
D05	-0.008	0.610
NetIncdif	0.195	0.000
Convexity	-1.179	0.019
Constant	-3.135	0.000
Observations	142,127	

Note: The table shows the Maximum Likelihood estimates of the coefficients of the structural probit equation for the choice between self-employment and employment using pooled data from 1999 to 2005. The dependent variable is binary with a value of 1 if the individual is self-employed and 0 if employed. The standard errors are bootstrapped with 1,000 replications. The data are annual cross-sections from the Survey of Labour and Income Dynamics (SLID).

Table 8: Estimates of the Structural Probit Equation, by Year

	1999	2000	2001	2002	2003	2004	2005
NetIncdif	0.474	0.292	0.159	0.650	0.353	0.242	0.005
	(0.001)	(0.006)	(0.157)	(0.000)	(0.003)	(0.074)	(0.969)
Convexity	-1.973	-2.032	-4.271	-0.239	-0.386	-1.840	-3.046
	(0.074)	(0.195)	(0.009)	(0.658)	(0.791)	(0.214)	(0.141)
Observations	22,009	20,515	19,832	20,680	20,047	18,677	20,367

Note: The table shows the Maximum Likelihood estimates of the structural probit equation for each year. The dependent variable is binary with a value of 1 if the individual is self-employed and 0 if employed. The results are displayed only for NetIncdiff and Convexity. The standard errors are bootstrapped with 1,000 replications. P-values are in parentheses. The data are annual cross-sections from the Survey of Labour and Income Dynamics (SLID).

Table 9: Federal Personal Income Tax Schedules, 2000 Tax Code and Flat Tax

	Tax Rate	
Tax Bracket	2000 Actual	Flat Tax
y ≤ $30,004	17%	20%
$30,004 < y < $60,009	25%	20%
y ≥ $60,009	29%	20%

Note: The second column of the table shows the federal personal marginal income tax rates in Canada in 2000 and the third column shows a hypothetical single rate tax reform.

Table 10: Comparisons of the Probability of Self-Employment (%) in 2000 for the Actual Tax Code and a Hypothetical Flat Tax Reform

Tax Scenario	Average	White Collar	Service	Blue Collar	Higher Income	Lower Income
2000 Tax Code	5.758	5.122	6.271	6.477	7.258	5.636
Flat Tax	5.803	5.194	6.284	6.497	7.462	5.667
Convexity: 2000 NetIncdif: Flat Tax	5.736	5.099	6.278	6.437	7.271	5.611
NetIncdif: 2000 Convexity: Flat Tax	5.825	5.218	6.278	6.538	7.449	5.692

Note: The table shows the predicted values of the probability of self-employment in year 2000 for selected categories of individuals. The average probability shown in the second column is for the entire sample of individuals in 2000. The subsequent columns show the probabilities for selected categories of individuals. Higher income refers to predicted incomes greater than $50,000. Lower income refers to predicted incomes less than $50,000. The first two rows of figures show the probabilities under the 2000 tax code and under a 20% flat tax, respectively. The third and fourth rows of figures show the partial effects of the flat tax reform by fixing either Convexity or NetIncdiff at its values under the 2000 tax code and the other variable at its values under the flat tax.