TAX PROGRESSIVITY AND SELF-EMPLOYMENT DYNAMICS

Wiji Arulampalam and Andrea Papini*

Abstract—Analysis of the relationship between taxes and self-employment should account for the interplay between responses in self-employment and wage employment. To this end, we estimate a two-state multispell duration model which accounts for both observed and unobserved heterogeneity using a large longitudinal administrative data set for Norway for 1993 to 2011. Our findings confirm theoretical predictions and are robust to various changes to definitions and sample selections. A policy experiment simulating a flatter tax schedule in the year 2000 is found to encourage self-employment, delivering a net increase of predicted inflow into self-employment from 2.8% to 5.3%.

I. Introduction

MODELS of choices facing wage earners typically neglect the fact that taxpayers may exit or enter self-employment because of differences in tax schedules. Since the interplay between the occupational choices is typically not considered in models of labor supply, these models are silent on how tax differences across occupational choice affect decisions.[1,2] However, in contrast, models of choice of the self-employed are dominated by perspectives where decisions are based on implicit or explicit comparisons to the wage sectors. One obvious reason for this asymmetry is the relative sizes of the sectors. For example, the self-employment rate (as a percentage of total employment) in Norway is 7%, whereas the European Union average is approximately 15% (OECD, 2018).

The relationship to the wage sector is not the only factor that complicates the assessment of the effects of taxation on self-employment. From a theoretical perspective, the tax effects are ambiguous. On the one hand, an increase in the tax rate may diminish the self-employment rate as it reduces expected returns. On the other hand, high taxes may encourage self-employment if loss offsetting is allowed because the government provides an implicit insurance by sharing the risk associated with self-employment (Domar & Musgrave, 1944).[3]

A large majority of empirical studies on the effect of taxes on the level of self-employment activity focuses on the United States. These studies examine the extensive margin in occupational choice models (see Bruce, 2000, 2002; Gentry & Hubbard, 2000, 2004; Schuetze, 2000; Schuetze & Bruce, 2004; Cullen & Gordon, 2007; and Moore, 2004).[4] Studies for other countries include Hansson (2012) for Sweden; Fossen (2007, 2009) and Fossen and Steiner (2009) for Germany, and Wen and Gordon (2014) for Canada. Results from these studies are mixed. Results for the United States, for example, do not provide an unambiguous answer about the relationship between tax progressivity and self-employment. However, in other countries, tax progressivity is generally found to discourage self-employment.[5]

The representation of the tax schedule is important in any analysis of tax effects on self-employment. Some studies include measures of marginal and/or average taxes in a quasi-experimental or reduced-form analysis to investigate the effect of nonlinearities in taxes on entrepreneurship.[6] In other studies, authors have used measures of expected net-income differences and/or tax progressivity to capture the tax effects. For example, Gentry and Hubbard (2000, 2004) use the spread in the marginal (or average) tax rates faced by a self-employed individual at various levels of "success," where success is defined as the observed distribution of the three-year real wage growth for entrants into self-employment.

In two recent studies (Fossen, 2009; Wen & Gordon, 2014), authors derive the tax variables within a structural framework

[1]See Blundell and MaCurdy (1999), and Keane (2011) for reviews of the literature on labor supply.

[2]"Occupational choice" here means a choice between wage employment and self-employment.

[3]The role of loss offsetting is less clear in the presence of a progressive tax schedule. If the tax rate is an increasing function of taxable income, the savings made because of the loss offset are usually lower in magnitude than the taxes paid on profits (Gentry & Hubbard, 2000).

[4]See Hansson (2012), Gale and Brown (2013), and Clingingsmith and Shane (2016) for surveys on taxation and self-employment.

[5]A positive correlation between taxes and self-employment may also partly be attributed to the higher tax evasion or avoidance possibilities in self-employment relative to wage employment (see, for instance, Schuetze & Bruce, 2004). Our data do not allow us to address this issue. Recent tax evasion estimates for Norway show that around 14% of the business income is not reported (Nygård et al., 2019). This estimate is lower than typical estimates for the United States but close to what is found among the self-employed in Finland (Johansson, 2005) and Denmark (Kleven et al., 2011). Slemrod (2007) estimates that around 57% of U.S. nonfarm business income was not reported. The time and individual unobservable effects included in our model will partially mitigate this problem if the differential evasion possibilities are relatively constant over the time period under consideration. Another issue is the possibility of a tax-induced organizational shift. See Papini (2018) for a recent analysis of this issue. We treat a self-employed individual who decides to incorporate and, thus, decides to earn wages from the company, as a wage earner. We also include region fixed effects to partly control for this issue, as this organizational shift was more common in some regions and time periods (Papini, 2018).

[6]For example, Bruce (2002), and Gurley-Calvez and Bruce (2008) use expected marginal tax rates, or, alternatively, average tax rates to capture nonlinearities in the tax schedule. These authors do not include any measure of riskiness of income received.

where the decision making is based on the difference in expected utilities. Yet the two papers differ in many aspects and draw different conclusions. The use of different utility functions and assumptions regarding the pretax income distribution of the individual result in different variables that capture the effects of nonlinearities in the tax schedule. They also use different statistical models (logit versus probit).

Fossen (2009) models the transitions between wage and self-employment using data from the German Socio-Economic Panel (GSOEP) over the period 2002 to 2006 and a logit model in which agents are assumed to trade-off risks and returns. He uses a constant relative risk-aversion utility and assumes normally distributed pretax income. The two relevant model-generated variables are (i) the difference in net-of-tax incomes in the two occupations and (ii) the variances of the individual's posttax income distributions in the transition equation.

In contrast, Wen and Gordon (2014) use a pooled cross-sectional sample from the Canadian Survey of Labour and Income Dynamics over the years 1999 to 2005 to estimate the probability of self-employment in a probit model.[7] They assume risk neutrality and a log-normal distribution for the pretax income. The relevant "tax variables" are (i) the difference in log net-of-tax incomes in the occupations (*netincdiff*) and (ii) a variable that they call *convexity*. The variable *convexity* has an intuitive interpretation as the "increase in tax-liability taken on by the self-employed due to the volatility of their business income, expressed as a proportion of their disposable income."

Both studies use selectivity-corrected income equations to predict individual pretax incomes and then use a tax-transfer microsimulation model to generate the relevant expected value and variance of after-tax incomes in wage employment and self-employment. The estimated models are subsequently used to simulate the effects of hypothetical tax policy scenarios that reduced progressivity. Fossen finds the "flatter-tax" reforms considered discourage individuals from choosing self-employment;[8] Wen and Gordon find a "small" positive effect on the probability of finding someone in self-employment.[9]

Here we use the two variables *netincdiff* and *convexity* used by Wen and Gordon (2014). Although some of the tax effects in both studies are captured via net-income differences, the additional variable *convexity* in Wen and Gordon (2014) is an individual-specific measure that intuitively captures the interaction between the progressivity of the tax schedule and

[7]Thus, the focus is on being in self-employment at the time of interview and not on entering self-employment.

[8]The interpretation given in Fossen (2009) is that a flatter tax schedule increases expected returns in self-employment, but at the same time it also increases the risk because the variance of the net income distribution also increases. The second effect is found to dominate the first one, and hence, a flatter tax schedule discourages self-employment.

[9]The "flatter-tax" reform considered is found to increase the probability of finding someone in self-employment by 0.04 percentage points from the base model prediction of 5.76%.

the volatility of self-employment income relative to wage income.

Our work complements the existing empirical literature in various ways. First, our definitions of wage employment and self-employment are based on reported incomes from tax records and not on survey responses. We use data drawn from various Norwegian population registers over the period 1993 to 2011. The data include rich sociodemographic information together with highly accurate income measures from the annual tax returns. Second, we model the evolution of employment spells using a two-state multispell duration model that controls for observed and unobserved heterogeneity correlated across spells and accounts for left and right censoring in the observed spells. This contrasts with several previous contributions, which mainly focus on self-employment entries or exits using survey data with self-reported employment status and short panels of individuals.

We generally find significant effects of both *netincdiff* and *convexity* on the probability of exit from both types of employment spells, conforming to theoretical predictions as discussed in section VA. The increase in *convexity* is found to increase the probability of exiting self-employment and to decrease the probability of entry into self-employment; that is, *convexity* has a discouraging effect on self-employment, ceteris paribus. On the other hand, an opposite effect is found for *netincdiff*: negative (positive) in the self-employment (wage-employment) equation. Additionally, in our base model, we find a larger effect of *convexity* relative to that of *netincdiff*, implying that small increases in *convexity* will require large increases in *netincdiff* to discourage the self-employed from quitting and to encourage wage earners to enter self-employment.

Given the way the tax variables are constructed, a change in the progressivity of the tax schedule will have an impact on the *convexity* and on the *netincdiff* by changing the expected net income difference in self-employment and wage employment. From this, the total effect on the rate of self-employment of a decrease in the progressivity of the tax schedule is hard to predict. Hence, to better understand the net effect, we simulate a tax experiment that replaces the personal income tax structure in the year 2000 with a less progressive, revenue-neutral tax schedule, as explained in section VB. The overall estimated effect of this policy change is positive on the share of self-employment. The average exit rate from self-employment is estimated to go down by 0.018 (s.e. 0.184) percentage points, and the estimated exit rate from wage-employment is estimated to increase by 0.119 (s.e. 0.010) percentage points. This change results in a net increase of predicted inflow into self-employment changing from about 2.8% to 5.3%.
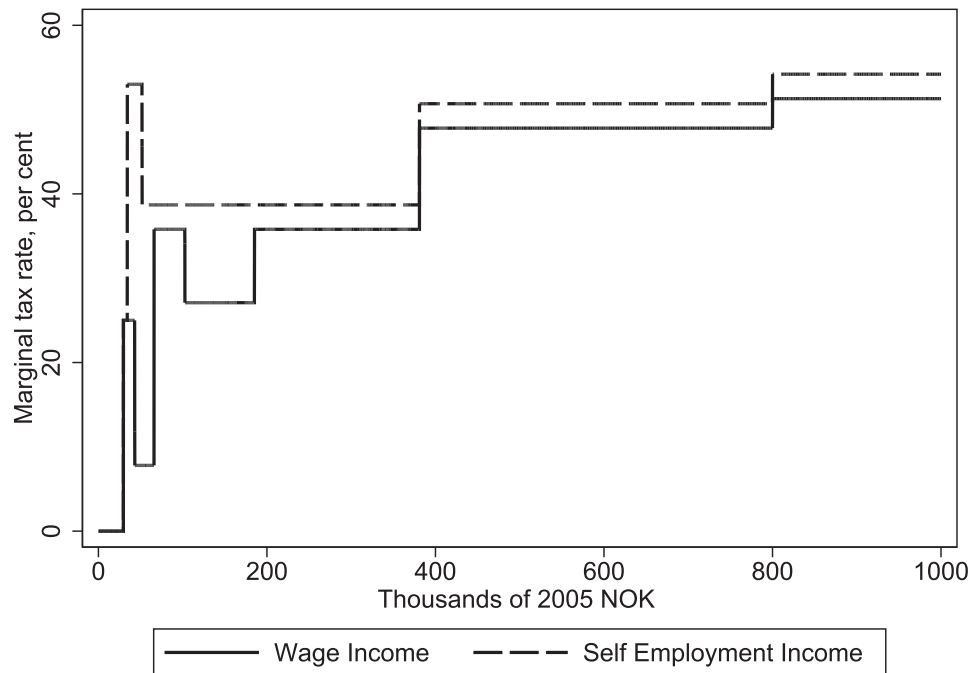
The rest of the paper is organized as follows. Section II describes the taxation of self-employment income and wages during our sample period. Section III sets out our econometric model. In section IV we provide details of the data and the sample selected for our analyses. We also present the procedure used for estimating the tax variables. The estimation

FIGURE 1.—MARGINAL TAX RATE FOR WAGE AND SELF-EMPLOYMENT INCOMES, YEAR 2005

(i) Solid line: Marginal tax rate for a wage earner in tax class 1 (see online appendix A.2 for the definition of tax class 1) with only wage income. Employer's social security contributions are excluded. (ii) Dashed line: Marginal tax rate for a self-employed individual in tax class 1 with only self-employed income and no capital invested in the firm.

results are discussed in section V along with the results from our policy simulation and some sensitivity checks. Finally, section VI concludes the paper.

II. Taxation in Norway

Tax reforms undertaken in 1992 introduced a dual-income tax system in Norway. Under this regime, all types of capital income are taxed at a flat rate, but a progressive schedule applies to labor and pension income. Individuals pay income tax on two different tax bases: (i) *ordinary* income and (ii) *personal* income.

Income from wages, self-employment, capital, transfers, and pensions are first grouped as *ordinary* income. After deductions, individuals pay tax at a flat rate (28% during most of the sample period) on *ordinary* income.[10] The other tax base—*personal* income—includes wage income, transfers, and pension income and self-employment income due to active efforts, but not capital income. Individuals pay a surtax and social security contributions levied on the *personal* income.

As an example, consider a wage earner whose only source of income is from wages in the year 2005. The solid line in figure 1 represents the marginal tax rates that apply to the wage income. No taxes and contributions are paid for income below the tax-free threshold. This threshold was NOK 29,600

in 2005.[11] Above the threshold, a social security contribution of 25% (of the *personal* income above NOK 29,600) is due, up to the amount where the total amount is the same as one would get using the standard rate of 7.8% on all *personal* income. Thereafter the rate is 7.8%. The flat tax on *ordinary* income (28% in 2005) is paid on the part of income that exceeds the sum of the personal allowance and the basic allowance. The basic allowance is 31% of wage income with a lower limit of NOK 31,800 and an upper limit of NOK 57,400. The personal allowance is a standard deduction from *ordinary* income, set at NOK 34,200 in 2005. The last two steps in figure 1 represent the two surtaxes that raise the marginal tax rates by 12 percentage points and 15.5 percentage points. The maximum marginal tax rate of 51.3% is reached after the two surtaxes become effective.

Taxation is more complicated for the self-employed because income represents the reward to the labor of the individual as well as the returns to the capital invested in the firm. Given the lower tax rate on capital income, the decision about how to declare the income was not left to the discretion of the self-employed; rules were established to split the profits into labor and capital income.[12] The dashed line in figure 1 represents the marginal tax rates that apply to self-employment income in the case where no capital is invested in the firm.

---

[10]The deductions include a standard personal allowance, a deduction for expenses including interest payments, and a basic allowance, which is a percentage (up to a maximum) of labor or pension income.

[11]The exchange rate in 2005 was 1 USD ≡ 6.45 NOK; 1 EUR ≡ 8.01 NOK.

[12]Capital income is calculated by multiplying the capital invested in the firm with a rate of return annually established by the government. The labor income is then estimated by subtracting the imputed capital income from the reported self-employment income net of expenses.

FIGURE 2.—MARGINAL TAX RATE FOR WAGE INCOME, YEARS 1995, 2005, AND 2010
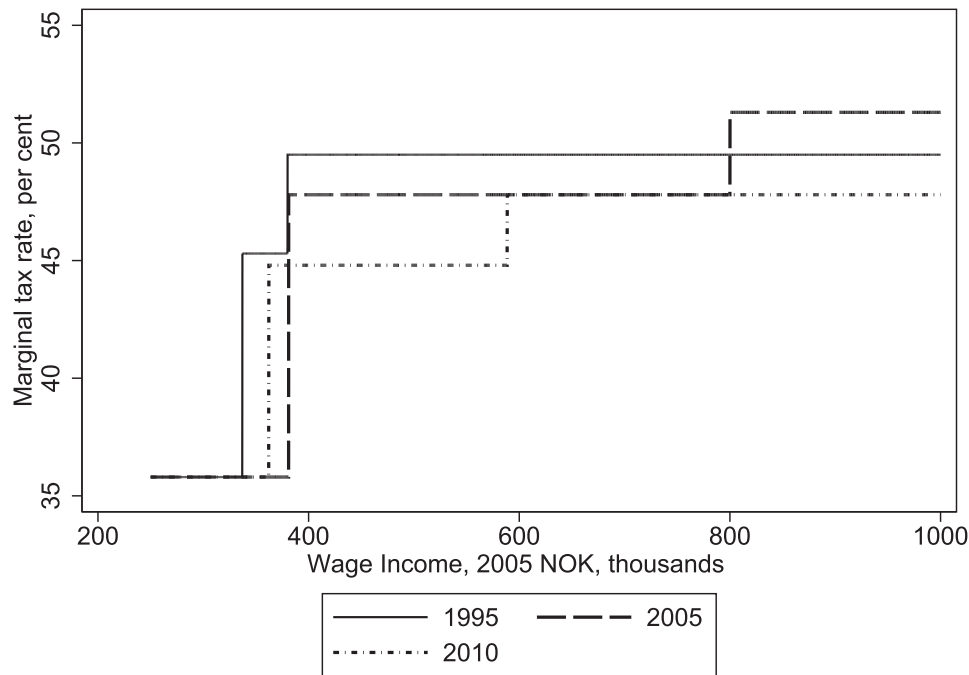
(i) Marginal tax rate for a wage earner in tax class 1 with only wage income in years 1995, 2005, and 2010. Employer's social security contributions are excluded. Thresholds are adjusted to account for income growth during the period (base year is 2005). Marginal tax rate is reported only for income larger than 200,000 NOK. (ii) To improve readability, the case for self-employment income is not reported, because it would imply only a proportional vertical shift of each of the three curves presented; see figure 1.

The main differences to the wage income case are the lack of basic allowance and the higher social security contribution (10.7% in 2005).

Tax progressivity is achieved through the tax-free allowances applied to ordinary income and the surtaxes on personal income. However, during the years under consideration, the progressivity changed several times because of changes to the tax rates, the number of surtaxes, and their thresholds. Overall, tax progressivity decreased during the period. Figures 2 and 3 show the marginal tax rates and average tax rates in different years for an individual whose only source of income was wage income.[13] Marginal tax rates in the year 2010 were overall lower than in the year 1995, and, for most part, they were also lower than in the year 2005. Similarly, the average tax rates in 1995 were in general higher than the rates in 2005 and 2010 (figure 3).

III. Econometric Model

Drawing heavily on the framework of Ham et al. (2016), we model employment transitions using a two-state multi-spell discrete duration model accounting for unobserved individual heterogeneity.[14] The two employment states are self-

employment and wage employment. The duration variable is measured in terms of the Norwegian financial year, which is the calendar year (January–December). Approximately 70% of individuals in our sample have a first spell that is left censored. Without dropping these individuals from the analysis sample, we include them and specify a different model of exit rates for them (Ham et al., 2016). We check for sensitivity of our estimates to excluding the left-censored spells, which is equivalent to using an inflow sample.

With regard to the unobserved heterogeneity, we follow the literature and assume this to be distributed independently across individuals and of the covariates included but fixed over the same type of spell, but correlated across the two employment states and the type of spell (fresh versus left-censored). A discrete distribution is assumed for the unobserved heterogeneity.

As we closely follow the setup in Ham et al. (2016), we provide only the form of the hazard function used and refer readers to their paper for further details. For notational simplicity, we do not distinguish between duration time and calendar time, although the estimated model does. The duration time random variable is denoted as Υ. Let $j = \{sf, wf, sc, wc\}$ where the first letter denotes a self-employment (s) or a wage-employment (w) spell, and the second letter denotes a fresh (f) or a left-censored spell (c). The probability that individual i would leave the spell in spell type j at the *end* of duration

---

[13]Note that the thresholds account for wage growth.

[14]Following the early pioneering work by Lancaster (1979), and Nickell (1979), the literature on modeling durations using survival analysis has developed very fast. Lancaster (1990) and Van den Berg (2001) provide a comprehensive discussion of theoretical issues as well as empirical examples that helped to develop this literature. See Carrasco and García-Pérez

(2015) for another recent application of a two-state multispell duration model with discretely distributed unobserved heterogeneity.

FIGURE 3.—AVERAGE TAX RATE FOR WAGE INCOME, YEARS 1995, 2005, AND 2010
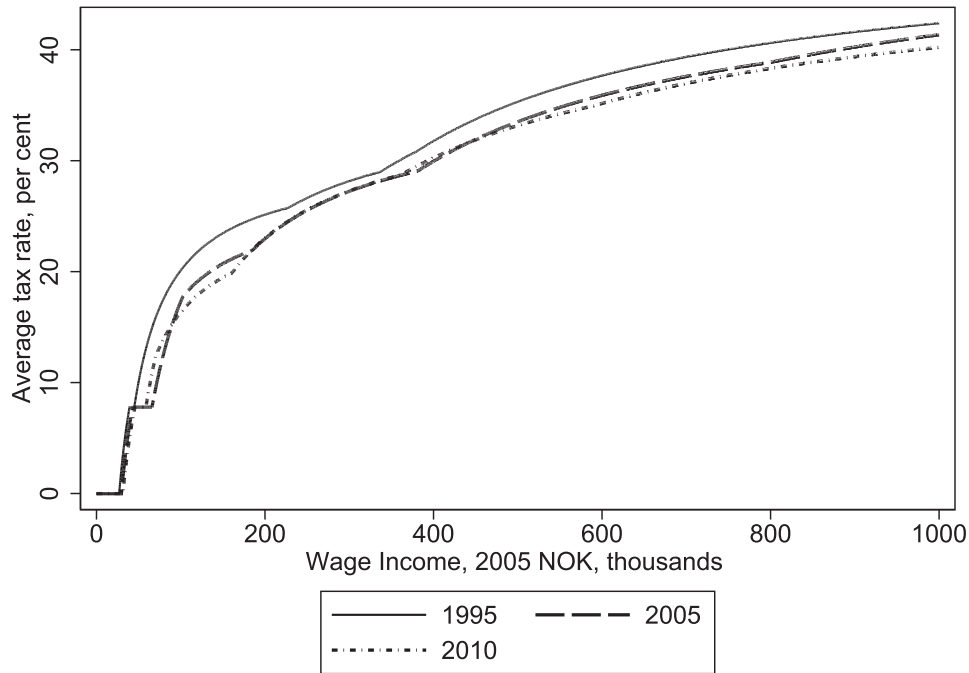
(i) Average tax rate for a wage earner in tax class 1 with only wage income in years 1995, 2005, and 2010. Employer's social security contribution are excluded. Thresholds are adjusted to account for income growth during the period (base year is 2005). (ii) To improve readability, the case for self-employment income is not reported, because it would imply only a proportional vertical shift of each of the three curves presented; see figure 1.

time t, conditional on not having left in $t-1$, is a discrete time hazard $\lambda(t)$ given by

$$\lambda_{i,j}(t|x_{i,j}, \omega_{i,j})$$
$$= \Pr(\Upsilon_{i,j} = t | \Upsilon_{i,j} > t - 1, taxation_{i,j}(t), x_{i,j}(t), \omega_{i,j})$$
$$= F\left(h_j(t) + x_{i,j}(t)'\beta_j + \alpha_j' taxation_{i,j}(t) + \omega_{i,j}\right), \quad (1)$$

where h_j is the duration dependence function, $x_{i,j}(t)$ contains time-fixed and time-varying observed individual characteristics, *taxation* contains the tax variable(s), and $\omega_{i,j}$ is the unobserved heterogeneity. F is specified as the complementary log-log distribution function.[15] To achieve convergence with stable parameter estimates, we restrict the duration dependence function to a log linear form and model the unobserved heterogeneity to be discrete with two points of support.[16] We keep the hazard-specific intercepts, set

[15]The distribution function is given by $F(z) = 1 - \exp[-\exp(z)]$. Some other popular distributions used are the standard normal and the logistic cdfs, which are symmetric distributions. The distribution we employ is not a symmetric distribution. A discrete time hazard model derived from an underlying continuous time proportional hazard model can be written in this form. See Narendranathan and Stewart (1993) for an application.

[16]Theoretical results exist for lack of nonparametric identification in hazard models when one or more of the following are present: duration dependence, time-varying variables, time-varying effects, and unobserved heterogeneity. For example, Baker and Melino (2000), using simulations, look at the behavior of the nonparametric maximum likelihood estimator for a discrete duration model with unobserved heterogeneity and unknown duration effect, and find the estimator to be biased when both are nonparametrically specified. Unsurprisingly, empirical researchers have also found the model

$\omega_{sf} = \omega_{ef} = \omega_{sc} = \omega_{ec} = 0$ as a normalization, and estimate the associated probability, p.

IV.   Data, Sample, and Variable Definitions

### A.   Data and Sample Selection

The present study benefits from rich longitudinal Norwegian administrative data for the period 1993 to 2011. The main data source is the *Income and Wealth Statistics for Persons and Families* (Statistics Norway, 2005). The data are drawn from the annual tax returns and the education registers (years of education and fields of studies). The data also contain individual and family sociodemographic characteristics. Our focus is on wage earners and the self-employed who have strong labor market attachment, and so we restrict our analysis to Norwegian citizens aged 25 to 61 and exclude those who have reported any income from agricultural, forestry, or fishing activities.[17]

We use an income-based definition to identify periods or spells of self-employment and wage employment. In our main analysis, we classify an individual observation as "self-employed" if the major source of income is self-employment

estimations to be unstable when most of the time effects are modeled in an unrestricted manner and have thus imposed some functional form restrictions to identify the parameters. See Ham and Rea (1987) for a discussion of these issues in the context of an empirical application.

[17]Since immigrants are a group of "selected" individuals, we exclude them.

FIGURE 4.—ANNUAL SHARE OF SELF-EMPLOYMENT OBSERVATION
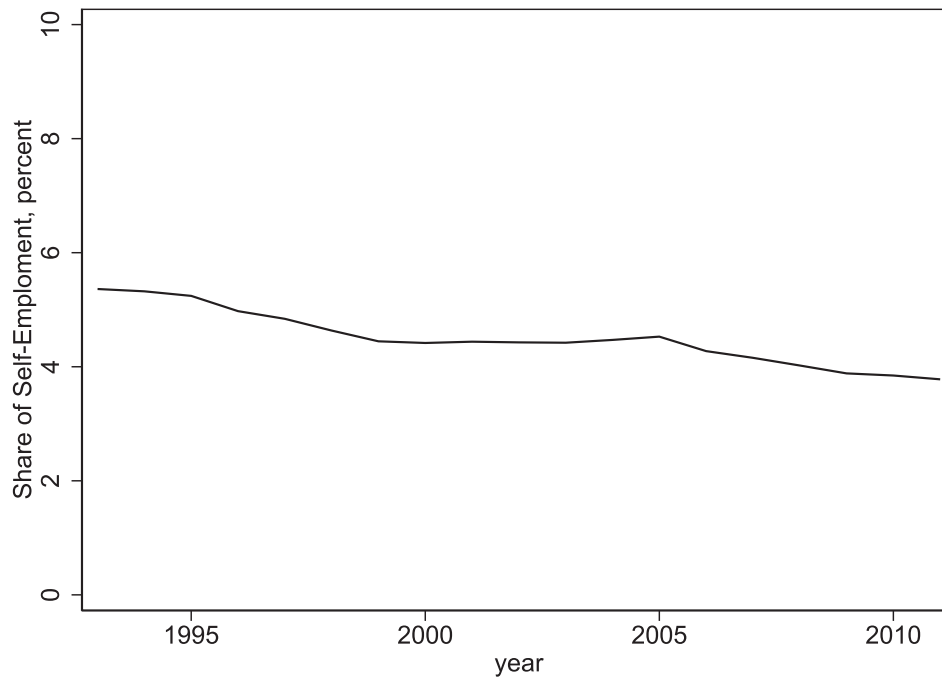
Annual self-employment observation as a share of total self-employment plus wage employment observations. Categorisation into self-employment and wage employment is described in section III.

income, i.e., if the reported self-employment income (net of expenses) is larger in absolute value than the wage income and is also larger than government transfers (which include disability insurance, unemployment benefits, and other types of pensions).[18] Additionally, we restrict our sample to those who have been classified as either being in wage employment or self-employment during the observation period 1993 to 2011.[19]

The majority of individuals never experience any self-employment spells. For example, the average rate of self-employment over the sample period is around 5% (see figure 4). To reduce the computational burden of working with more than 2 million individuals, we use a 50% random sample to generate our tax variables. From this sample, we next randomly select 2% of individuals who have never been categorized as self-employed and 20% from the other group, which includes individuals with periods of self-employment spells only and individuals with a mix of types of employment. This gives us a sample of 476,275 individual-year unweighted observations. All analyses presented use sample weights to account for this endogenous sample selection, following Solon, Haider, and Wooldridge (2015).

B.   Defining and Estimating the Tax Variables

Our analysis is based on the theoretical exposition of an expected utility maximization approach discussed by Wen and Gordon (2014), who in turn base their model on the one developed by Rees and Shah (1986). Assuming risk neutrality, a convex tax schedule, and log-normally distributed pretax income, they show how the probability of self-employment can be written as a function of the tax schedule using two representations of the effects of taxation.[20] These are (1) *net-incdiff,* which is the difference in log of expected net incomes in self-employment and wage employment, and (2) *convexity*, which is a measure of how the expected tax liability changes due to the volatility of their self-employment income relative to the net income in wage employment (see online appendix A.1 for further details).

The construction of the two tax variables requires net-income distributions for each individual. We use a tax simulator to generate these (see online appendix A.2). The simulator considers the yearly rules for taxing self-employment income net of expenses, wages, and other sources of income. Other sources of income are taken to be exogenous; these are added to the predicted self-employment or wage income. The simulator also accounts for the main deductions and allowances, as well as for the system for taxation of the labor and capital parts of net self-employment income; see section II.

---

[18]We also exclude individuals who do not report any wage income or business income that is larger than the "Basic amount" during the observation period for at least three years. The "Basic amount" is the base for calculating many of the Norwegian social insurance scheme's payments and was 78,024 NOK in 2011 (the approximate exchange rate in that year was: 1 USD ≡ 5.67 NOK; 1 EUR ≡ 7.79 NOK).

[19]Around 18% of the individuals in the sample experienced at least one "third-state" spell (periods of time that cannot be defined either as wage employment or as self-employment) and are omitted from the analysis.

[20]Wen and Gordon (2014) represent the convex tax function specifying the after-tax income x_j as $(y_j)^{1-\tau} y_0{}^\tau$, where the tax parameters τ and y_0 are such that $0 < \tau < 1$, and $y_0 > 0$ represents the income at which the tax liability is zero. $(1 - \tau)$ is the elasticity of posttax income with respect to pretax income (also see Musgrave & Thin [1948] and Benabou [2000]).

TABLE 1.—SUMMARY STATISTICS: MEAN (STD DEVIATION)

	All	WE Sample	SE Sample
Individual-specific variables			
Females	0.47 (0.50)	0.48 (0.50)	0.27 (0.44)
Lower secondary school and less	0.39 (0.49)	0.35 (0.49)	0.53 (0.50)
Upper secondary school	0.30 (0.46)	0.31 (0.46)	0.27 (0.45)
University	0.32 (0.47)	0.34 (0.47)	0.20 (0.40)
Time-varying variables			
Age at the start of the spell	35.06 (9.24)	34.84 (9.20)	39.80 (8.80)
Years 1993–1998	0.30 (0.49)	0.30 (0.46)	0.34 (0.47)
Years 1999–2002	0.22 (0.41)	0.22 (0.41)	0.21 (0.41)
Years 2003–2007	0.27 (0.44)	0.27 (0.44)	0.27 (0.44)
Years 2008–2011	0.21 (0.41)	0.21 (0.41)	0.18 (0.39)
Eastern Norway	0.50 (0.50)	0.49 (0.50)	0.55 (0.50)
Southern Norway	0.05 (0.22)	0.05 (0.22)	0.06 (0.24)
Western Norway	0.26 (0.44)	0.26 (0.44)	0.24 (0.42)
Central Norway	0.09 (0.28)	0.09 (0.29)	0.07 (0.26)
Northern Norway	0.10 (0.30)	0.10 (0.30)	0.08 (0.27)
Local unemployment rate	2.73 (0.83)	2.73 (0.83)	2.78 (0.83)
convexity	0.007 (0.008)	0.007 (0.008)	0.012 (0.008)
netincdiff	−0.448 (0.19)	−0.429 (0.17)	−0.825 (0.25)
Proportion of exits		0.006	0.106

(i) Years covered in the analysis are 1993–2011. (ii) Definitions of wage employment and self-employment and the sample selection criteria used are provided in section IV. (iii) All averages and proportions are based on the weighted sample (see section IV for further details). (iv) The number of unweighted observations is 476,275, of which 362,217 are classified as wage employment and 114,058 as self-employment. (v) The number of unweighted individuals is 34,746.

Our construction of the two tax variables closely follows Wen and Gordon (2014). Assuming pretax income to be log-normally distributed, $y_j \sim LN(\mu_j, \sigma_j)$, where $j = s$ for self-employment, and $j = e$ for wage employment, we have

$$\bar{y}_j \equiv E(y_j) = \exp\left(\mu_j + \frac{1}{2}\sigma_j^2\right). \tag{2}$$

The first tax variable, *netincdiff*, that enters the occupational choice probability is given by

$$netincdiff = [(1 - \tau_s)\ln(\bar{y}_s)]/[(1 - \tau_e)\ln(\bar{y}_e)]$$
$$\simeq \ln[netincome_s/netincome_e], \tag{3}$$

where τ is a tax parameter from the tax function (see footnote 21). For each individual, we first estimate the selectivity-corrected *expected* pretax income ($\bar{y}_j$) for each occupation in each period.[21] We then use the tax simulator to generate the individual specific net incomes in both occupations: $netincome_s$ and $netincome_e$.

Next, we define the second individual specific tax variable representation: *convexity*. This variable is defined as the difference between the expected tax liability $E[T(y_s)]$ and the tax liability at the expected income $T(\bar{y}_s)$, relative to the expected net income $\bar{x}_s = (\bar{y}_s - T(\bar{y}_s))$.[22] Wage employment is generally less riskier than self-employment. Hence, following Wen and Gordon (2014), we derive our *convexity* variable by setting the coefficient of variation for wage income equal

to 0, so that *convexity* is associated with uncertainties in self-employment income only.

The *convexity* variable for each individual in each time period is calculated as

$$convexity = \frac{E[T(y_s)] - T(\bar{y}_s)}{\bar{y}_s - T(\bar{y}_s)}. \tag{4}$$

C. Summary Statistics

Summary statistics for the main estimation sample are provided in table 1. On average, in the weighted sample, the proportion of individuals exiting out of a period of work and into a period of self-employment is less than 1%, whereas the average share of exits out of a period of self-employment is 11%. We next turn to our tax variables.

The overall distributions of the two tax variables are provided in figures 5 and 6. *netincdiff* is predominantly negative, indicating that, for the majority of observations in the sample, the predicted net wage income is higher than the predicted net self-employment income.[23] *convexity* is as expected, estimated to be mostly positive.[24] The average value of predicted *netincdiff* of −0.448 implies that the *net* income in self-employment is about 64% of net income in wage employment. The average estimated value of *convexity* is 0.007

---

[21] Online appendix A.3 contains the full set of estimates from the equations that we used to generate the income variables.

[22] As shown in Wen and Gordon (2014), the tax liability function $T(y_j)$ in the theoretical model is given by $y_j(1 - (y_0/y_j)^\tau)$. This term is strictly convex and hence the use of the term *convexity*; see Wen and Gordon (2014, p. 472).

[23] The paradox of self-employment being characterized by higher uncertainty and lower earnings than wage employment is a common finding in previous studies (see, for example, Hamilton [2000] and Hurst & Pugsley [2011], or Berglann et al. [2011] for the case of Norway). There are several possible explanations for this puzzle. Among them, (i) the relevance of unobserved nonpecuniary benefits; (ii) unobserved underreporting of income by the self-employed; and (iii) overestimation by the self-employed of their probability of success.

[24] Negative *convexity* values are possible if the tax function is not convex. Estimated *convexity* is 0 for about 1.5% of the observations and negative for about 5.5% of the observations.

FIGURE 5.—DENSITY OF *netincdiff*
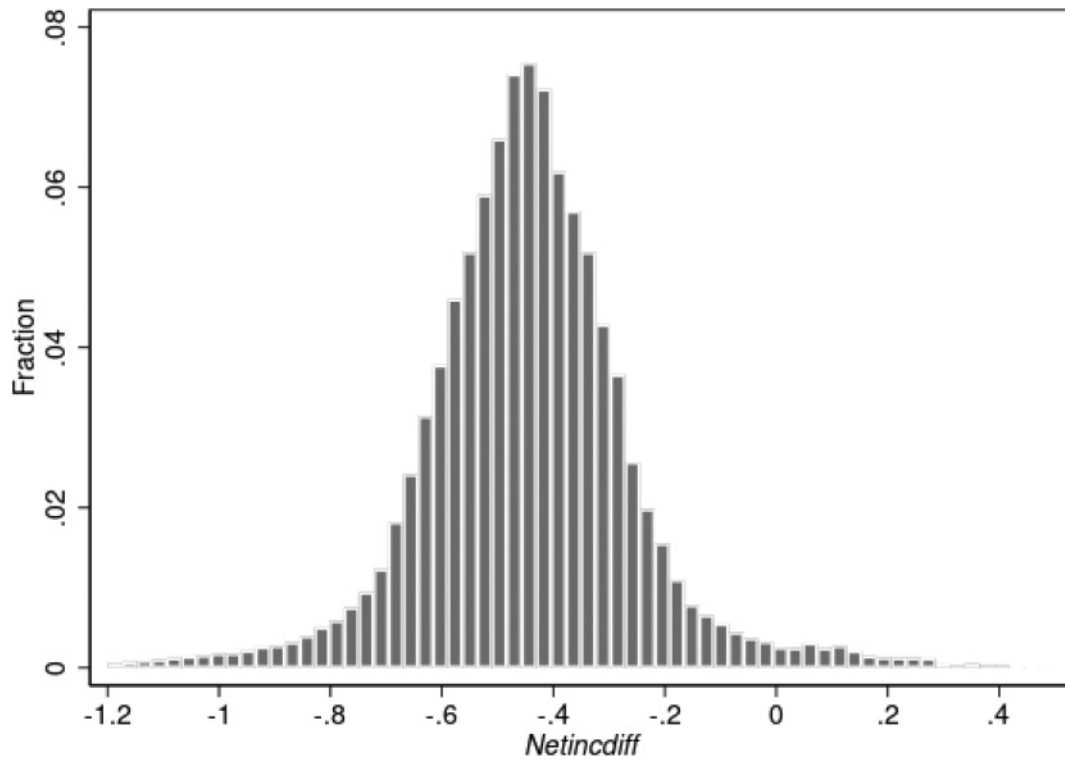
netincdiff distribution across all years and observations. *netincdiff* is defined in equation (3).

FIGURE 6.—DENSITY OF *convexity*
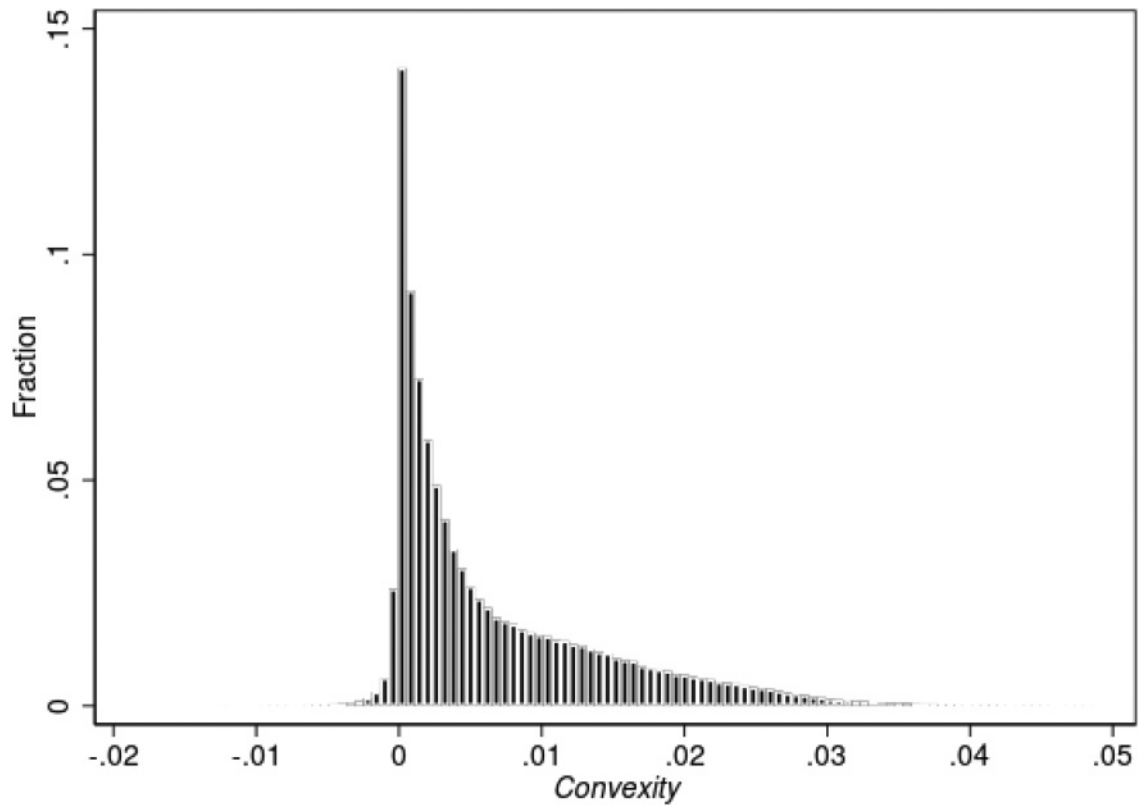
convexity distribution across all years and observations. *convexity* defined in equation (4).

FIGURE 7.—BOX-AND-WHISKER PLOT FOR *netincdiff*

(i) *netincdiff* = ln[net income in *SE*/net income in *WE*]. See section IVB for further details. (ii) The box shows the median and the interquartile range (IQR). The end of the whiskers gives 1.5 times IQR.

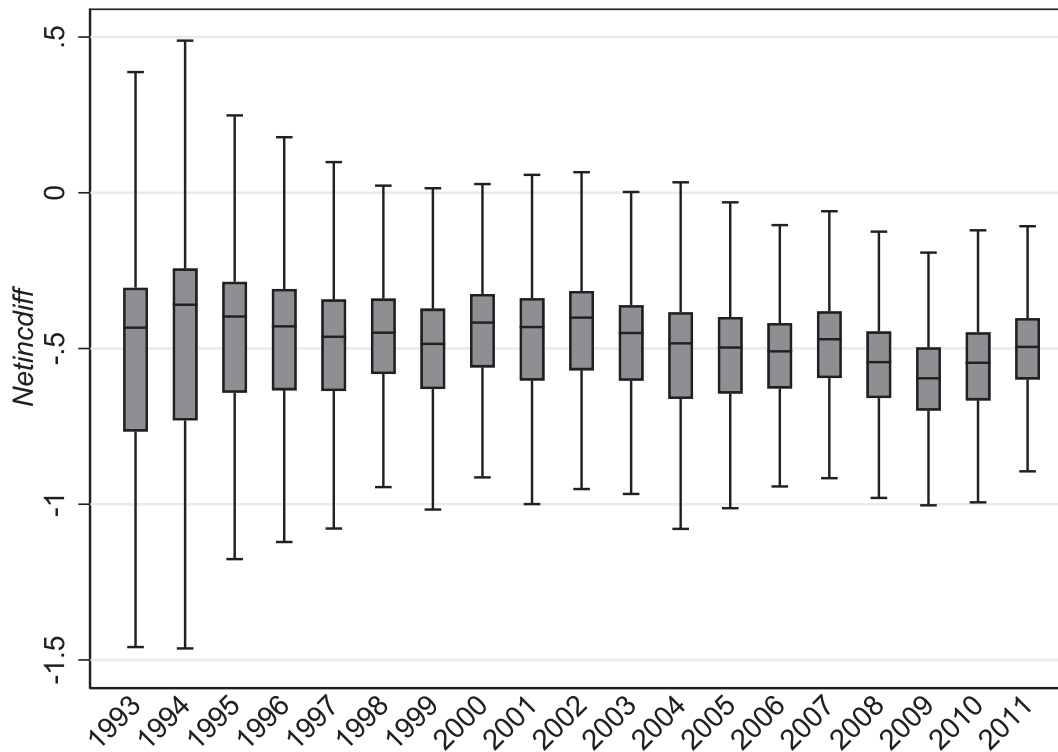(s.d.= 0.008), which is lower than the *convexity* value of 0.011 (s.d. 0.16) reported by Wen and Gordon (2014) for Canada.

Box-and-whisker plots in figures 7 and 8 show how these estimated tax variables change over time. The median *netincdiff* remains stable over time without experiencing a clear trend, and the spread decreases over time. A slightly declining trend is observed for *convexity*, which complies with the reduced progressivity of the taxation during the sample period (section II). The temporary up-tick in the median and spread of *convexity* in 2000 is consistent with the fact that two surtaxes were introduced in that year, making the overall tax schedule more progressive.[25,26]

---

[25] Another possible explanation for this is the increased uncertainty due to the recession in the early 2000s.

[26] We carried out an analysis of covariance to assess the contribution of various factors to the variation of the two tax variables. We included all the variables (sex, marital status, education, region, children, family head, year dummies, two selection correction terms, and estimated variances) that were used in the predictions of these two tax variables along with the "other" tax variable (*convexity* or *netincdiff*). The model R-squared values were 29% and 49% respectively in the *netincdiff* and convexity equations. The top four largest contributors explained 46% of the model sum of squares (SS) in the *netincdiff* equation. These were education, selection into *SE*, and the regional and year dummies. With regard to the convexity variable, the top four largest contributors were the year effects, education, and estimated heteroskedastic functions, which together explained 38% of the model SS. The convexity (*netincdiff*) variable in the *netincdiff* (*convexity*) equation explained less than 4% (2%) of the model variations. The largest contributions to the model SS came from the year effects.

In addition to the two tax variables, the models also include time-varying and time-invariant control variables. The time-invariant variables are sex, age at the start of the spell, indicator variables for highest education level achieved, and regional dummies to account for local labor market conditions. Calendar time dummies control for macroeffects. The data are an unbalanced panel; see descriptive information in table 1. Self-employed individuals are on average older and less educated than individuals who are paid wages, and a lower proportion of females is found among the self-employed. Self-employment is also highly concentrated in the more densely populated areas of eastern Norway (the Oslo region) and western Norway (the Bergen region).

V. Results

A. Main Results

Before discussing the parametric model estimation results, we provide a plot of the empirical hazard in figure 9.[27] The raw data self-employment (*SE*) hazard consistently lies above the wage-employment (*WE*) hazard, implying that the conditional exit rate from *SE* is higher relative to an exit from *WE*. The *WE* hazard is quite low and stable over the spell duration. The probability of exiting from *SE* into *WE* is around

---

[27] This is the number of individuals exiting during the year divided by the number of individuals in that state at the beginning of the year.

FIGURE 8.—BOX-AND-WHISKER PLOT FOR *convexity*
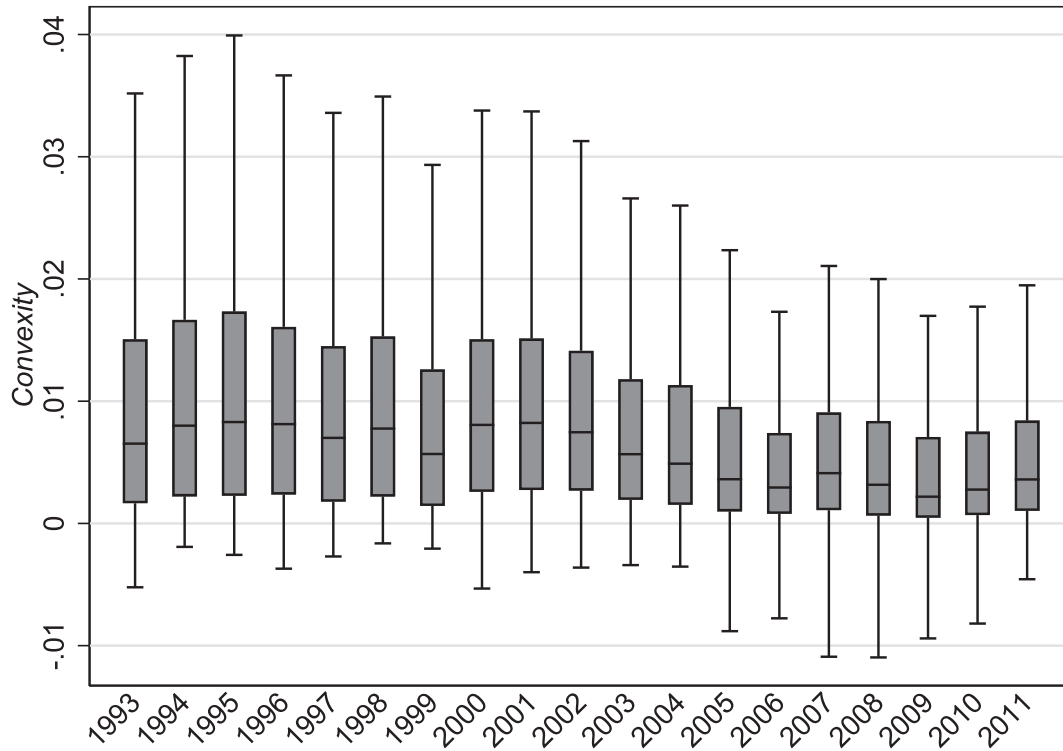
(i) See equation (4) for the definition of *convexity*. (ii) The box shows the median and the interquartile range (IQR). The end of the whiskers gives 1.5 times IQR.

FIGURE 9.—NONPARAMETRIC HAZARD ESTIMATES
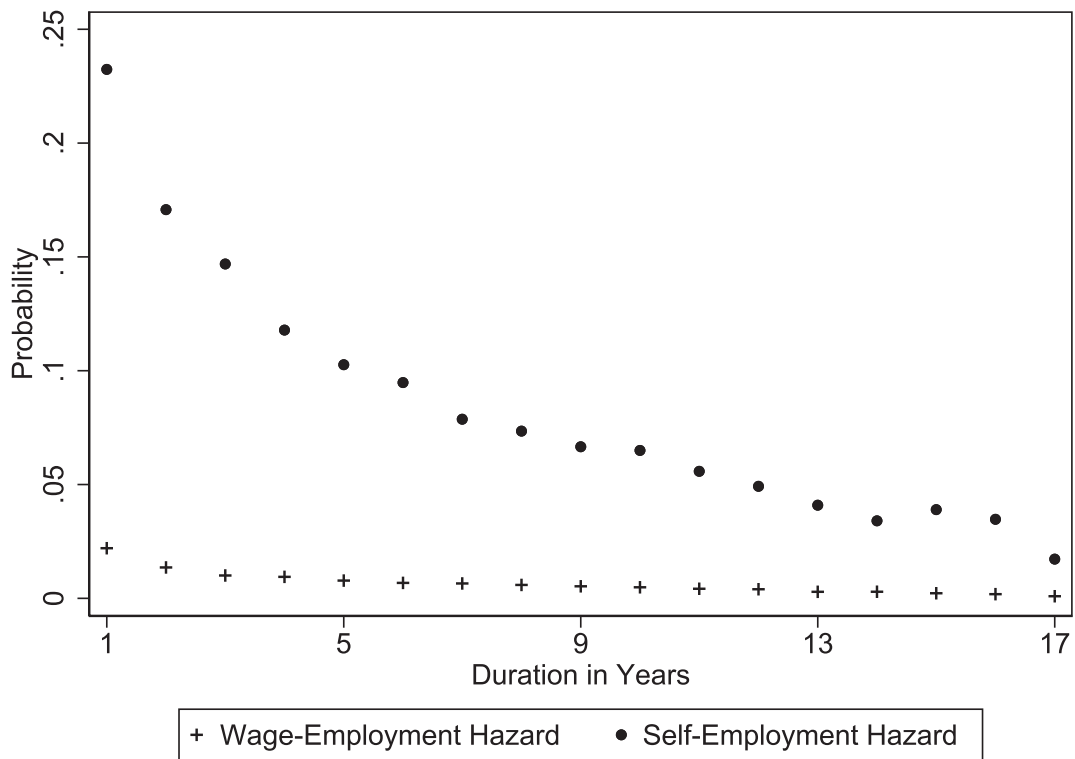
The figure presents the nonparametric hazard estimates for wage employment and self-employment spells. These are the OLS estimated coefficients on the duration time dummies in a linear regression of the duration variable. The duration variable takes the value of 0 if that particular year refers to an ongoing spell and 1 when it is associated with an exit.

TABLE 2.—HAZARD MODEL ESTIMATES, MAIN SAMPLE

	Fresh Spells		Left-Censored Spells	
	SE [1]	WE [2]	SE [3]	WE [4]
netincdiff	−0.429 (0.053)	1.685 (0.082)	−0.725 (0.109)	1.753 (0.087)
convexity × 100	0.049 (0.015)	−0.246 (0.021)	−0.017 (0.030)	−0.163 (0.023)
Male	−0.024 (0.027)	0.602 (0.030)	0.191 (0.058)	0.776 (0.037)
Age at the start of the spell	−0.012 (0.001)	0.030 (0.002)	−0.034 (0.002)	−0.046 (0.002)
High school	−0.006 (0.029)	0.115 (0.035)	−0.131 (0.048)	−0.008 (0.038)
University	0.220 (0.028)	0.131 (0.037)	0.051 (0.051)	0.100 (0.038)
ln(duration)	−0.520 (0.016)	−0.490 (0.018)	−0.016 (0.037)	−0.234 (0.032)
Constant	−1.135 (0.092)	−3.11 (0.103)	−0.892 (0.192)	−1.930 (0.118)
Support points	−0.531 (0.049)	−3.042 (0.200)	−1.337 (0.072)	−1.839 (0.094)
Probability masses				
p1 (constants + support points)	0.805 (0.019)			
p2 (constants only)	0.195 (0.019)			
N observations (unweighted)	476,275			
N individuals (unweighted)	34,746			
Maximized log likelihood value	−105,687.67			

(i) MLE standard errors in parentheses. (ii) The models are estimated using a random sample of individuals as detailed in section IV. (iii) Omitted education category is no-education/high-school drop-out. (iv) The model additionally includes region and time indicators; see table 1. Complete sets of results are available in the online appendix A.4.

0.23 in the first year of the spell compared to 0.02 from WE into SE.

Our base model estimates are presented in table 2.[28] All four hazard functions are estimated simultaneously. Except for the left-censored SE hazard, the other three hazards show negative duration dependence, ceteris paribus. Insignificant duration dependence estimated for the left-censored SE spells is consistent with the observation that the probability of exiting is almost zero for high duration spells, and the sample of left-censored spells has a higher probability of containing large-duration spells.

We focus our discussions on the interpretation of the estimated effects of the tax variables. The theory predicts a positive (negative) effect of the netincdiff variable on the probability of exit from WE (SE). For example, the higher the proportionate increase in the net-income differential with respect to the net income from WE, the higher the exit rate from WE (Wen & Gordon, 2014; Taylor, 1996; Fossen, 2009). On the other hand, the theoretical prediction of the effect of convexity is negative on exit rate from WE since higher "convexity" would be expected to discourage SE. The estimated effects of the two tax variables conform to these theoretical predictions.

[28]The bootstrapped standard errors to account for the tax variables being "generated regressors" did not change the significance of our variables compared to the usual maximum likelihood standard errors for our base model reported in table 2. Hence, we report only the usual MLE standard errors in this table and subsequent tables.

These estimated coefficients are also found to be higher in absolute value for WE exit probabilities (columns 2 and 4). These results suggest that, compared to exits from SE, the probability of an exit from WE is more sensitive to changes in both expected net-income differences and tax progressivity. This is consistent with the fact that the SE tend to continue their business activities even if they experience lower earnings growth (Hamilton, 2000).

These estimates also indicate that a one percentage point increase in convexity requires an increase of approximately nine to fourteen percentage points in netincdiff to keep these hazards unchanged. Note that increases in convexity in this calculation are assumed to take place via changes to the volatility of SE income (online appendix A.1 equation [A.4]) because we assume no uncertainty in WE income in the calculation of this variable. Similarly, the increase in netincdiff is assumed to work either via a reduction in the pretax income in WE or via an increase in the expected pretax SE income (not altering the variance of the SE income distribution). To further explore these effects accounting for the relationship between the two tax variables, we simulated a policy experiment. The results are presented below.

B. Results from a Policy Experiment

So far we have looked at the effects of partial changes in the tax variables netincdiff and convexity. Motivated by the analysis in Wen and Gordon (2014), to gain further understanding

FIGURE 10.—MARGINAL TAX RATE FOR WAGE INCOME, YEAR 2000, AND HYPOTHETICAL UNIQUE SURTAX ON PERSONAL INCOME
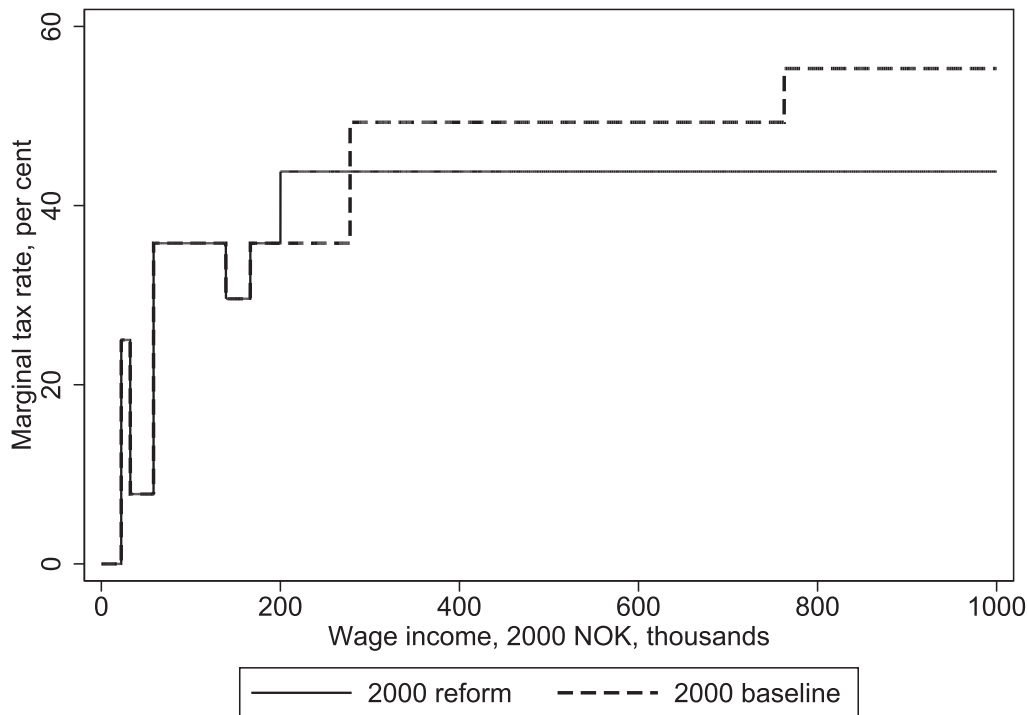
(i) Bold line: Marginal tax rate for a wage earner in tax class 1 (see online appendix A.2) with only wage income in year 2000. Employer's social security contributions are excluded. (ii) Dashed line: year 2000 tax experiment. The two surtaxes are replaced by a single surtax of 11% for gross incomes exceeding 200,000 NOK. (iii) To improve readability, the case for self-employment income is not reported, because it would only imply a proportional vertical shift of each marginal tax curve presented; see figure 1.

of how these related changes may be achieved through taxation, we consider a hypothetical reform in the year 2000. We chose this year because the Norwegian government introduced two changes in the taxation of gross income from wage and self-employment in that year. The threshold for the 1999 surtax rate of 13.5% was increased from 269,100 NOK to 277,800 NOK. More importantly, an additional surtax was introduced for income exceeding 762,700 NOK (dashed line in figure 10). These changes increased the overall progressivity of the Norwegian income tax system.[29]

Our policy experiment is to replace two of the surtaxes applied to personal income with one surtax, to create a flatter tax schedule (solid line in figure 10). The surtax value of 11% on gross income above 200,000 NOK is chosen to ensure revenue neutrality, given a "no behavioral reaction" assumption. Other features of the taxation are held constant. New values of *netincdiff* and *convexity* were generated under the hypothetical scenario using our tax simulator and the transition rates predicted from the estimated models.

The average values of the *netincdiff* and *convexity* variables in our weighted sample are −0.374 and 0.0071 under the new policy regime, compared to the original figures for the year 2000 of −0.382 and 0.0087, respectively. As expected, the less progressive tax schedule leads to a decrease of 0.16 percentage points in *convexity*. The hypothetical policy also

leads to a small increase in the mean *netincdiff*, so that average ratio of net income in *SE* to net income in *WE* changes from 68.2% to 68.8%.

The predicted transition probabilities and the corresponding standard errors, under the old and the new tax regimes, are reported in table 3.[30] In the benchmark year 2000, the model predicts that around 9.33% of self-employed individuals will transit out of *SE* to *WE* (case A).[31] However, the reform reduces this figure to 9.32% (case B). Under the new regime, the predicted transitions from *WE* to *SE* are higher at 0.68% compared to 0.56% in the base model. Since a very large proportion of individuals are in *WE* compared to *SE*, even with this small increase in the exit rates out of *WE* can generate a substantial net inflow into *SE*. The change in the exit rates induced by the policy reform is not significant for the self-employed.

To further explore how the model predicts responses to separate changes in the two tax variables, we look at these effects separately. In case C we hold the *convexity* variable fixed at a value that is the same as in the base case scenario and let the *netincdiff* variable change. Conversely, in case D we see a change in the *convexity* variable only. Table 3 shows that the partial effect of a change in *netincdiff* is an increase in transitions out of both *SE* and *WE*. This result is consistent

---

[29] According to exchange rates for 2000: 1 EUR ≡ 8.11 Norwegian kroner (NOK), and 1 USD ≡ 8.81 NOK.

[30] All predictions including the differences in predicted exit rate, and the associated standard errors, use all four hazards. These are calculated using STATA's *margins command*.

[31] The observed exit rates in 2000 were 9.813% and 0.595%.

TABLE 3.—AVERAGE PREDICTED EXIT PROBABILITIES (%) UNDER THE TAX REFORM SCENARIO

		Probability of Exit	
Case	Tax Scenario	from *SE*, %	from *WE*, %
A	Base model: year 2000, two surtaxes	9.334	0.562
	(s.e.)	(0.227)	(0.011)
B	Reform scenario: year 2000, one surtax	9.316	0.682
	(s.e.)	(0.289)	(0.016)
	Change A − B	0.018	−0.119
	(s.e.)	(0.184)	(0.010)
	Sample size in year 2000	6,043	130,019
C	*convexity*: unchanged from baseline *netincdiff*: reform	9.622	0.571
	(s.e.)	(0.234)	(0.011)
D	*netincdiff*: unchanged from baseline *convexity*: reform	9.034	0.673
	(s.e.)	(0.276)	(0.015)

(i) Actual exit rates in 2000 were 9.813% and 0.595%. (ii) Predicted exits are based on the estimated model from table 2. (iii) The percentage exits are calculated with respect to the stocks in each of the occupational categories. (iv) Case A refers to the actual situation as it was in year 2000 with two surtaxes; Calculated *convexity* and *netincdiff* in this scenario were used in the estimation of the main model. (v) Case B refers to a hypothetical reform scenario that replaces two surtaxes with just one surtax. New values of *convexity* and *netincdiff* are recalculated given the new tax rules. (vi) Case C considers values of *convexity* from the baseline scenario and values of *netincdiff* from the reform scenario. (vii) Case D considers values of *netincdiff* from the baseline scenario and values of *convexity* from the reform scenario. (viii) The above predictions and the associated standard errors were calculated using the delta method in STATA's command *margins*. Average exit rates as well as the differenced average exit rates were all calculated using all four hazards. (ix) All calculations are based on the weighted sample.

with the fact that mean *netincdiff* experiences a decrease in the reform scenario for the self-employed, whereas it increases for wage earners. A possible explanation for this effect is that the reduced progressivity of the tax system would encourage a larger share of wage earners who expect to be successful in self-employment to transit into *SE*. On the other hand, because a majority of self-employed individuals have been predicted to have a higher posttax income in regular employment, a flatter tax scenario would increase the proportion of them leaving *SE* for *WE*. In contrast, the decrease in *convexity*, common to both *WE* and *SE* observations, reduces the transitions from *SE* and increases the exit from *WE*. In summary, the hypothetical tax scenario is found to encourage the net inflow into *SE*. Translating these estimates to numbers, we find that such a policy would have resulted in an increase from 2.76% to 5.34% in the net inflow into *SE*.[32]

Finally, we briefly compare our results to the findings of Wen and Gordon (2014), given that the same variables are used to capture the effects of taxes and uncertainty. Wen and Gordon (2014) also simulated the effect of a flatter tax schedule in the year 2000 using Canadian data. Their policy reform implied decreases in the average values of (i) *netincdiff* and *convexity* from −22.5% to −23.3% (a decrease of 4%) and (ii) from 1.2% to 0.8% (a reduction of 33%). The policy reform we considered increased the average values of *netincdiff* by around 2%, and reduced the average values of *convexity* by 18%. From the simulated policy reform, Wen and Gordon (2014) estimate an increase in the number of self-employed individuals of 0.78% (5.76% to 5.80%), which is substantially below our estimate of 2.6% (our experiment implies an increase of the self-employment share in 2001 from 4.56% to 4.68%). One should however note that Wen and Gordon (2014) do not model transitions.

---

[32]The predicted probability of exit from *SE* in the reform scenario is not statistically significantly different from the base model, and so we use the base model predicted probability. With the reform scenario prediction, the predicted net inflow would rise to 5.36%.

C. Sensitivity Checks

In this subsection we present results of some of our investigations into key assumptions of our empirical approach. We consider the following: (i) redefinition of a self-employment spell; (ii) estimation based only on the inflow sample; (iii) trimming the *netincdiff* with respect to extreme values; (iv) controlling for local unemployment rates; (v) including a dummy variable for individuals receiving some unemployment insurance during the year; and (vi) allowing for the share of capital in *SE* income to be nonzero. Table 4 reports the results of these investigations. The estimated effects of the tax variables are qualitatively unchanged. The full set of results is available in the online appendix A.3.

Our first investigation examines the influence of the definition of an *SE* spell. In our base model we included individuals in the sample if they had at least three years of labor market attachment, that is, if the net *SE* income or *WE* is larger in absolute value than the *basic amount* for at least three years over the years the individual is observed in data. We now redefine the sample requiring only one year of labor market attachment. The results using this new definition are presented in panel B of table 4. Individuals with less attachment to the labor market would be expected to be more sensitive to changes in the tax variables, and this is what we find when we include these individuals in the estimation sample. The results are qualitatively similar to the results from our base case (panel A). However, the coefficient for *convexity* in the *SE* fresh spells hazard decreased substantially. Individuals with less attachment to the labor market with low predicted *SE* income might be expected to be less sensitive to the progressivity of the tax system.

The base model was estimated using both the left-censored and fresh spells. We reestimate our model using only the inflow sample. This reduces the total number of unweighted observations to 229,036. The definition of an *SE* spell is the same as the one used in our base model. The results are presented in panel C of table 4. The results are broadly similar

TABLE 4.—SENSITIVITY CHECKS: HAZARD MODEL ESTIMATES

Variables	Fresh Spells		Left-Censored Spells	
	SE [1]	WE [2]	SE [3]	WE [4]
A. Base case				
netincdiff	−0.429	1.685	−0.725	1.753
	(0.053)	(0.082)	(0.109)	(0.087)
convexity × 100	0.049	−0.246	−0.017	−0.163
	(0.015)	(0.021)	(0.030)	(0.023)
B. Changes to sample definition				
netincdiff	−0.493	1.734	−0.615	1.768
	(0.016)	(0.026)	(0.034)	(0.028)
convexity × 100	0.011	−0.277	0.016	−0.187
	(0.005)	(0.007)	(0.009)	(0.007)
C. Excluding left-censored spells				
netincdiff	−0.405	1.920		
	(0.053)	(0.083)		
convexity × 100	0.055	−0.292		
	(0.015)	(0.022)		
D. Using trimmed netincdiff				
netincdiff	−0.333	2.281	−0.871	2.998
	(0.068)	(0.108)	(0.138)	(0.134)
convexity × 100	0.065	−0.222	−0.061	−0.237
	(0.017)	(0.025)	(0.032)	(0.026)
E. Including regional unemployment rate 1996–2011				
netincdiff	−0.531	1.709	−0.718	1.568
	(0.057)	(0.096)	(0.110)	(0.083)
convexity × 100	0.045	−0.292	0.047	−0.115
	(0.018)	(0.026)	(0.029)	(0.025)
F. Including regional dummies 1996–2011				
netincdiff	−0.519	1.762	−0.754	1.607
	(0.057)	(0.095)	(0.110)	(0.081)
convexity × 100	0.036	−0.314	0.038	−0.140
	(0.018)	(0.026)	(0.030)	(0.024)
G. Including unemployment benefits dummy				
netincdiff	−0.415	1.698	−0.694	1.763
	(0.053)	(0.082)	(0.109)	(0.087)
convexity × 100	0.049	−0.252	−0.014	−0.165
	(0.015)	(0.022)	(0.030)	(0.023)
H. Using 3.7% capital income invested in SE				
netincdiff	−0.434	1.712	−0.719	1.761
	(0.052)	(0.083)	(0.108)	(0.086)
convexity × 100	0.058	−0.264	−0.008	−0.166
	(0.016)	(0.024)	(0.031)	(0.025)

(i) Standard errors in parentheses. (ii) See section VC for further details. (iii) Panels E and F report results with no unobserved heterogeneity (see footnote 35). (iv) Also see notes to table 2. (v) A full set of results are available in the online appendix A.3.

to our base model results. As expected, dropping those spells for which we have no information about the length of time they had spent in a particular state prior to the sample start slightly increases the estimates.

The third investigation involves omitting observations with extreme predicted values for the variable *netincdiff*. As shown in figure 5, the distribution of *netincdiff* exhibits some lumpiness in the tails. To assess the effect of extreme values of *netincdiff*, we drop those individuals who have at least one occupation-specific *netincdiff* above the top 1% or below the 1% cut-off values.[33] Since individuals with very high or low *netincdiff* would be expected to be less sensitive than the oth-

ers, we would expect the estimated effects of *netincdiff* to be higher in absolute values. This is what we see with the results reported in panel D. In the base model (panel A), we found the *WE* exits to be more sensitive than the *SE* exits, and now we see that the effect of *netincdiff* goes up for the *WE* exits without much change for in *SE* exits.

The next investigation examines the influence of local labor market conditions. In the main specification we use regional dummies to partially control for labor market conditions. Perhaps a better control for local labor market conditions would be the use of local unemployment rates. Unfortunately such information is available only from 1996, so we report two sets of results. In panel E we substitute the regional dummies with regional unemployment rates. In panel F we reestimate

[33]To preserve a continuous series of observations, all observations belonging to an individual are dropped if we find at least one *netincdiff* that is either less than the first percentile or above the 99th percentile value for that individual resulting in a loss of more than 2% of the sample. We lose about 9% of the observations, resulting in 432,409 observations in our un-

weighted sample. The definition of a *SE* spell is the same as the one used in our base model.

our base model using the restricted sample of 1996 to 2011. The results are very similar to each other and qualitatively similar to the baseline results.[34]

As described in section IV, in our base model we drop individuals who received more in social security benefits than their self-employment income or wages in any year. However, it can be the case that individuals are unemployed for a short period and the unemployment insurance is small enough so that the individual is still defined as a self-employed or a wage earner. Individuals with an interruption to their work might behave differently from individuals transiting directly from *WE* to *SE*. We therefore include a dummy variable for those individuals who received unemployment insurance during the year. As panel G shows, the results are similar to those from the base model.

In Norway self-employed individuals have the option of having a share of the self-employed income declared as capital income, which is taxed at a lower rate than labor income, as explained in section 2. Tax variables used in our main model are generated under the assumption that the share of capital income in total income is zero (see online appendix A.2). We believe our assumption is reasonable for the following reasons. First, it is not clear what is an appropriate assumption regarding the proportion of capital income used in the generation of counterfactual *SE* income distributions for the wage earners, which are also exogenous. Second, during our sample period, the share declared as capital income is either 0 or very small (median value is 0.037). However, we check for sensitivity by regenerating our tax variables allowing for 3.7% of the predicted *SE* income to be reported as capital income instead of 0. The results are given in panel H. The effect of convexity is slightly stronger on the *SE* exit rates, and the rest of the estimated effects remain similar to the base model estimates.

VI. Conclusion

We look at the effect of taxation on self-employment and wage employment durations. Our work complements the existing literature on many dimensions. First, in contrast to many existing studies, our definitions of self-employment and wage employment are based on income reported in Norwegian tax returns. The rest of the variables used come from various other registry data. Norwegian registry data are considered to be exceptional in terms of coverage and reliability (Blundell, Graber & Mogstad, 2015). Second, we look at the evolution of self-employment and wage employment spells over a very long period, from 1993 to 2011. We model these transitions using a two-state multispell duration model allowing for correlated unobserved heterogeneity and controlling for a rich set of sociodemographic characteristics.

We focus on the effects of two tax variables: *netincdiff* and *convexity*, obtained from Wen and Gordon (2014). *netincdiff* is defined as the difference in log net-of-tax income in the two occupations, and *convexity* is an individual-specific measure that captures the interaction between the progressivity of the tax schedule and the volatility of self-employment income relative to wage income. We use the model to predict the transitions under a simulated tax regime that reduced the progressivity of the tax schedule in the year 2000. We also provide some sensitivity checks with respect to the definition of self-employment, the selection of the estimation sample, and other factors. The estimated effects of our two tax variables of interest are qualitatively unchanged, and the quantitative differences are as expected.

The main finding is that, as predicted by theory, higher expected net earnings in self-employment relative to wage employment reduces the probability of exiting out of a self-employment spell. The entry into self-employment—or equivalently the exit out of wage employment—is found to be more sensitive to changes in the two variables than exit from self-employment. In our base model, the estimated effect of changes to *netincdiff* that are required when *convexity* changes by a percentage point, to encourage self-employment, is about nine to fourteen times larger in percentage point terms. To shed further light on this, we carried out a policy experiment by implementing a flatter tax schedule in the year 2000 that resulted in reduced tax progressivity. The hypothetical scenario was found to encourage entry into self-employment but not significantly the exit from self-employment, with the estimated inflow into self-employment increasing to 5.34% from the base model prediction of 2.76%.

REFERENCES

Baker, Michael, and Angelo Melino, "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics* 96:2 (June 2000), 357–393. 10.1016/S0304-4076(99)00064-0

Benabou, Roland, "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review* 90:1 (March 2000), 96–129. 10.1257/aer.90.1.96

Berglann, Helge, Espen R. Moen, Knut Røed, and Jens Fredrik Skogstrøm, "Entrepreneurship: Origins and Returns," *Labour Economics* 18:2 (April 2011), 180–193. 10.1016/j.labeco.2010.10.002

Blundell, Richard, and Thomas MaCurdy, "Labor Supply: A Review of Alternative Approaches" (pp. 1559–1695), in Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics*, vol. 3 (Amsterdam: Elsevier, 1999). 10.1016/S1573-4463(99)03008-4

Blundell, Richard, Michael Graber, and Magne Mogstad, "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family," *Journal of Public Economics* 127 (2015), 58–73. 10.1016/j.jpubeco.2014.04.011

Bruce, Donald, "Effects of the United States Tax System on Transitions into Self-employment," *Labour Economics* 7:5 (September 2000), 545–574. 10.1016/S0927-5371(00)00013-0

—— "Taxes and Entrepreneurial Endurance: Evidence from the Self-employed," *National Tax Journal* 55:1 (March 2002), 5–24. 10.17310/ntj.2002.1.01

Carrasco, Raquel, and J. Ignacio García-Pérez, "Employment Dynamics of Immigrants Versus Natives: Evidence from the Boom-Bust Period in Spain, 2000–2011," *Economic Inquiry* 53:2 (April 2015), 1038–1060. 10.1111/ecin.12175

Clingingsmith, David, and Scott Shane, "How Individual Income Tax Policy Affects Entrepreneurship," *Fordham Law Review* 84:6 (2016), 2495–2516.

[34]We made multiple attempts but were unable to find significant unobserved heterogeneity in these models with the reduced number of years. We therefore report results from the model where we set the unobserved heterogeneity component to 0.

Cullen, Julie Berry, and Roger H. Gordon, "Taxes and Entrepreneurial Risk-Taking: Theory and Evidence for the U.S.," *Journal of Public Economics* 91:78 (2007), 1479–1505. 10.1016/j.jpubeco.2006.12.001

den Berg, Gerard J. Van, "Duration Models: Specification, Identification and Multiple Durations" (pp. 3381–3460), in J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, vol. 5 (Amsterdam: Elsevier, 2001). 10.1016/S1573-4412(01)05008-5

Domar, Evsey D., and Richard A. Musgrave, "Proportional Income Taxation and Risk-Taking," *Quarterly Journal of Economics* 58:3 (1944), 388–422. 10.2307/1882847

Fossen, Frank M., "Risky Earnings, Taxation and Entrepreneurial Choice: A Microeconometric Model for Germany," Discussion Papers of DIW Berlin 705, DIW Berlin, German Institute for Economic Research (2007).

——— "Would a Flat-Rate Tax Stimulate Entrepreneurship in Germany? A Behavioural Microsimulation Analysis Allowing for Risk," *Fiscal Studies* 30:2 (2009), 179–218. 10.1111/j.1475-5890.2009.00093.x

Fossen, Frank M., and Viktor Steiner, "Income Taxes and Entrepreneurial Choice: Empirical Evidence from Two German Natural Experiments," *Empirical Economics* 36:3 (June 2009), 487–513. 10.1007/s00181-008-0208-z

Gale, William, and Samuel Brown, "Small Business, Innovation, and Tax Policy: A Review," *National Tax Journal* 66:4 (December 2013), 871–892.

Gentry, William M., and R. Glenn Hubbard, "Tax Policy and Entrepreneurial Entry," *American Economic Review* 90:2 (May 2000), 283–287. 10.1257/aer.90.2.283

——— "'Success Taxes,' Entrepreneurial Entry, and Innovation," NBER working paper 10551, National Bureau of Economic Research, Cambridge, MA (June 2004).

Gurley-Calvez, Tami, and Donald Bruce, "Do Tax Cuts Promote Entrepreneurial Longevity?," *National Tax Journal* 61:2 (June 2008), 225–250. 10.17310/ntj.2008.2.04

Ham, John C., and Samuel A. Rea Jr., "Unemployment Insurance and Male Unemployment Duration in Canada," *Journal of Labor Economics* 5:3 (July 1987), 325–353. 10.1086/298150

Ham, John C., Xianghong Li, and Lara D. Shore-Sheppard, "The Employment Dynamics of Disadvantaged Women: Evidence from the SIPP," *Journal of Labor Economics* 34:4 (2016), 899–944. 10.1086/686274

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-employment," *Journal of Political Economy* 108:3 (June 2000), 604–631. 10.1086/262131

Hansson, Åsa, "Tax Policy and Entrepreneurship: Empirical Evidence from Sweden," *Small Business Economics* 38:4 (May 2012), 495–513. 10.1007/s11187-010-9282-7

Hurst, Erik, and Benjamin Wild Pugsley, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity* 43:2 (Fall 2011), 73–142. 10.1353/eca.2011.0017

Johansson, Edvard, "An Estimate of Self-employment Income Underreporting in Finland," *Nordic Journal of Political Economy* 31 (2005), 99–109.

Keane, Michael P., "Labor Supply and Taxes: A Survey," *Journal of Economic Literature* 49:4 (December 2011), 961–1075. 10.1257/jel.49.4.961

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark," *Econometrica* 79:3 (2011), 651–692. 10.3982/ECTA9113

Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica* 47:4 (July 1979), 939–956. 10.2307/1914140

——— *The Econometric Analysis of Transition Data*, Econometric Society Monographs (Cambridge, UK: Cambridge University Press, 1990).

Moore, Kevin B., "The Effects of the 1986 and 1993 Tax Reforms on Self-employment," Finance and Economics Discussion Series 2004–05, Board of Governors of the Federal Reserve System (U.S.) (2004).

Musgrave, Richard A., and Tun Thin, "Income Tax Progression, 1929–48," *Journal of Political Economy* 56:6 (1948), 498–514. 10.1086/256742

Narendranathan, Wiji, and Mark B. Stewart, "How Does the Benefit Effect Vary as Unemployment Spells Lengthen?," *Journal of Applied Econometrics* 8:4 (1993), 361–381. 10.1002/jae.3950080405

Nickell, Stephen J., "Estimating the Probability of Leaving Unemployment," *Econometrica* 47:5 (September 1979), 1249–1266. 10.2307/1911961

Nygård, Odd E., Joel Slemrod, and Thor O. Thoresen, "Distributional Implications of Joint Tax Evasion," *Economic Journal* 129:620 (2019), 1894–1923.

OECD, *OECD Labour Force Statistics 2017* (Paris: OECD Publishing, 2018), available at http://dx.doi.org/10.1787/oecd_lfs-2017-en.

Papini, Andrea, "Tax Incentives and Choice of Organisational Form of Small Businesses. Identification through a Differentiated Payroll Tax Schedule," ISER working paper 2018-07 (June 2018).

Rees, Hedley, and Anup Shah, "An Empirical Analysis of Self-employment in the U.K.," *Journal of Applied Econometrics* 1:1 (1986), 95–108. 10.1002/jae.3950010107

Schuetze, Herb J., "Taxes, Economic Conditions and Recent Trends in Male Self-employment: A Canada-US Comparison," *Labour Economics* 7:5 (2000), 507–544. 10.1016/S0927-5371(00)00012-9

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

Slemrod, Joel, "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives* 21:1 (March 2007), 25–48. 10.1257/jep.21.1.25

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge, "What Are We Weighting for?," *Journal of Human Resources* 50:2 (2015), 301–316. 10.3368/jhr.50.2.301

Statistics Norway, "Income Statistics for Persons and Families 2002–2003," NOS Publication no. D338, available at http://www.ssb.no/en/inntekt-og-forbruk/statistikker/inntpf?fane=arkiv#content (September 2005).

Taylor, Mark P., "Earnings, Independence or Unemployment: Why Become Self-employed?," *Oxford Bulletin of Economics and Statistics* 58:2 (1996), 253–266. 10.1111/j.1468-0084.1996.mp58002003.x

Wen, Jean-François, and Daniel V. Gordon, "An Empirical Model of Tax Convexity and Self-employment," *Review of Economics and Statistics* 96:3 (July 2014), 471–482. 10.1162/REST_a_00388