Centre for
Economic and
Business Research

CEBR

Discussion Paper 2007-04

# Earnings, Uncertainty, and the Self-Employment Choice

Daniel le Maire
Bertel Schjerning

2007-04

# EARNINGS, UNCERTAINTY, AND THE SELF-EMPLOYMENT CHOICE

DANIEL LE MAIRE AND BERTEL SCHJERNING

Abstract. This paper investigates the relationship between self-employment choice, expected earnings, and uncertainty. Several interesting results emerge from our analysis on Danish longitudinal register data: Firstly, self-employed (taxable) personal income bunch at kink points in the tax system since self-employed can retain earnings and thereby transfer income across tax-years. Secondly, expected income level and income variance are important determinants in choice of occupation. Thirdly, men put more emphasis on expected earnings level, while women appears more risk averse, which contribute to explain why fewer women are self-employed. Finally, our results suggest that non-western immigrants are marginalized into self-employment.

Keywords: Occupational choice, self-employment, wage-differentials, income uncertainty, risk aversion, overconfidence, self-selection, gender differences.

JEL codes: J16, J24, J31, C33, C35.

---

*Date*: January 11, 2007.

We acknowledge the support from the Centre for Applied Microeconometrics (CAM). The activities of CAM are financed from a grant by the Danish National Research Foundation. Bertel Schjerning also acknowledges the support from Centre for Economic and Business Research (CEBR). The current project is financed by the The National Agency for Enterprise and Housing. None of our sponsors are responsible for the conclusions.

We gratefully acknowledge the comments we have received from Karsten Albæk, Martin Browning, Mette Ejrnæs, Steen Jørgensen, Ulrich Kaiser, Claire Kelly, Nikolaj Malchow-Møller, Jacob Roland Munch, Søren Leth-Petersen, Jacob Palsgaard Petersen, Michael Svarer and Lise Vesterlund. The usual disclaimer applies.

## 1. Introduction

Compared to wage work, self-employment is a fundamentally different occupation with respect to the type and source of income. While wage workers receive a wage which is subject to a relatively small level of uncertainty, self-employed individuals often face considerably more variation in their income. Moreover, since self-employed typically use own wealth to finance their business, they bear the risk associated with starting up the firm. Therefore, the expected income and the uncertainty of this income are likely to be important determinants of an individual's occupational choice.

The main objective of this paper is to investigate the relationship between the occupational choice and the distributions of associated monetary gains in different occupations. Specifically, we analyze how existing earnings differentials and differences in income uncertainty can explain observed occupational choices. A particular focus is on explaining why fewer women choose to become self-employed.

If individuals are risk averse, we would expect that the self-employed should be compensated for facing higher income uncertainty. However, earnings-differentials may arise for other reasons than risk compensation: Hamilton (2000) argues that cross-sectional earnings differentials may arise due to i) different earnings-experience profiles, ii) self-selection, and iii) non-pecuniary benefits. Hamilton finds that mean and median incomes are lower in self-employment than in wage-employment in the US, although those in the higher income brackets earn more in self-employment than in wage-employment. Hamilton concludes that individuals choose self-employment primarily because of non-pecuniary benefits.

An alternative (or complementary) explanation is that those who choose to become self-employed may be less risk-averse than the typical wage employed. They may even be risk-lovers. In a recent paper by Elston, Harrison, and Rutström (2006), experiments are used to characterize the attitudes to risk among entrepreneurs. Their main finding is that full-time entrepreneurs are less risk-averse and exhibit a significant joy of winning compared to non-entrepreneurs and part-time entrepreneurs.

Yet another explanation relates to the individual's subjective assessment of the probability of success. While Coelho and de Meza (2006) provide experimental evidence that entrepreneurs tend to overestimate their chance of success, Elston, Harrison, and Rutström (2006) do not find systematic judgmental error of profitability. However, it is found that part-time entrepreneurs are reluctant to enter markets where profitability is based on their perception of their relative skill ability.

Evidence from existing Danish questionnaire surveys shows that men focus more on the expected income level than women when choosing occupation, whereas women emphasize non-pecuniary benefits (Statistics Denmark, 1999; and Kjeldsen and Nielsen, 2000). Thus, 90 per cent of the women who had a child in the age of 0-2 years at the time of the business start-up state that an important reason for becoming self-employed was to make it easier to combine family life and work.

With respect to risk aversion, Byrnes, Miller, and Schafer (1999) analyze 150 psychological studies of risk-taking behavior, and find that in 14 out of 16 tasks, women are more risk-averse. However, according to Croson and Gneezy (2004) the evidence of women being more risk-averse is less clear in the economics literature which has typically focused on financial risk.

Several studies have suggested that overconfidence is part of human nature, e.g. Svenson (1980) reports that 90 per cent of Swedish drivers rate themselves above average. Recently and in relation to occupational choice, Niederle and Vesterlund (2006) find from the conduction of experiments that more women than men prefer to work under a non-competitive piece-rate compensation system rather than under a competitive tournament compensation scheme even though women are found to be as productive as men. Niederle and Vesterlund (2006) conclude that the reason for this difference is that men are too overconfident and enjoy competition more. In other words, too many low productivity men enter the competitive tournament, while productive women do not enter enough.

To evaluate whether the self-employed actually are compensated for their risk-taking, individual level information about the expected income (and the expected distribution of income) in both self-employment and wage-employment is required. To obtain this information, we estimate earnings functions for self-employed and wage-employed separately. However, individuals would be expected to select themselves into the type of occupation where they are most productive. Therefore, we estimate earnings functions for each occupational choice, using the dynamic panel data sample selection model of

Vella and Verbeek (1998, 1999). This also allows us to disentangle the role of unobserved heterogeneity and state dependence in the occupational choices. We find evidence of state dependence in the occupational choices.

The estimated earnings functions are then in turn used to predict an individual's income (and the uncertainty of this income) in different occupations. The random components of the model are partitioned into transitory and permanent shocks, which in turn are used to create occupational and education specific measures of income variance (uncertainty) and skewness (the risk/chance of very low/high incomes).

Rather that rather than characterizing the entrepreneur, we directly evaluate the impact of earnings on the choice of becoming self-employed, wage-employed or unemployed by examining the roles of expected earnings, risk aversion and over-confidence. This is done for each gender separately. Our results complement existing evidence from experimental economics, providing an potential explanations for the substantial gender gap in the probability of choosing to become self-employed.

We use a large longitudinal data set based on Danish register data from 1980 to 1996, providing us with detailed individual information about income, wealth, education, labor market status (occupation), region of residence, and immigration status. Since the panel covers more than 15 years, we can track long sequences of individual occupational choices and, thereby, appropriately investigate the dynamics of the self-employment choice.

Our results point to a large role for monetary aspects when choosing occupation. As expected, people prefer the sector with the highest expected income and lowest expected variance and, thus, on average appear risk-averse. We find that men put more emphasis on the earnings level, while women appear more risk-averse, which could be one of the crucial reasons why fewer women are self-employed. We do not find evidence of overconfidence. If anything, women instead seem to under-estimate their chance of success compared to men.

The explanatory power of the occupational choice model is quite impressive considering that we only include the predicted income level, variance and skewness. However, we explain much less of the variation in the realized occupational choices for the group of non-western immigrants. Immigrants are interesting with respect to occupational choice since they are more likely to start up their own business than natives. We find that immigrants put much less emphasis on the earnings level. These findings provide additional evidence for immigrants being marginalized into self-employment as Blume, Ejrnæs, Nielsen, and Würtz (2005) suggest. From their analysis on Danish transition data it is found that most non-western immigrants entering self-employment come from unemployment and that they do not use self-employment as a stepping stone for becoming wage-employed.

The rest of the paper is organized as follows: In section 2 we describe the data used in the analysis. In section 3 we formulate the econometric specification. In section 4, we present the results. Section 5 concludes.

## 2. Data

The data we use in this paper is an unbalanced panel data set for 1980-1996. The data is a representative 10 per cent sample extract drawn from the Integrated Database for labor market Research (IDA) and the Danish Income Registry (IKR) both maintained by Statistics Denmark. IDA and IKR are both longitudinal data based on register data for all individuals in Denmark. Since data originates from administrative records covering the entire Danish population there is only natural attrition in the data, i.e. birth, death and migration of individuals. The occupational status is observed once a year (the last week of November). We divide the labor market status into three states; self-employed, wage-employed, and unemployed. Since the panel covers more than 15 years, we have the possibility to track individuals over long time periods (before, during and after self-employment) and, thereby, appropriately control for the dynamics of the occupational choice. These high-quality Danish data contains very detailed individual information concerning, e.g., income, wealth, education, labor market status, region of residence, and immigration status. Moreover, the data also includes the same information for cohabitants allowing us to aggregate variables to the household level.

In order to avoid distortions in the results due to retirement patterns and educational attainment we restrict the sample to include persons aged 30-55 years only. This leaves us with $2,424,694$ observations in total of which $1,130,635$ are women.

For the analysis of occupational choice we need to decide on an income measure to use. One obvious candidate is disposable income since this measure is closely related to current consumption possibilities and, hence, utility.[1]

<< Figure 1 about here >>

<< Table 1 about here >>

Figure 1 shows kernel densities for the disposable income for self-employed and for wage-employed in 1996. Both distributions are right-skewed with the distribution of incomes from self-employed being most right skewed. From both Figure 1 and Table 1 it can be seen that the mean disposable income for self-employed is considerably below the mean income for wage-employed. However, due to the skewness the 90th percentile earns more in self-employment than the equivalent in wage-employment.

Figure 1 and Table 1 confirm the US evidence presented in Hamilton (2000), who also finds that mean and median incomes are lower in self-employment, but that those in the higher percentiles earn more in self-employment relative to wage-employment.

In Figure 2 we have depicted the (taxable) personal income for respectively wage-employed and self-employed together with two dotted vertical lines indicating where the medium and upper tax brackets set in. In contrast to wage-employed, self-employed tend to bunch just below where the tax brackets set in. This can be due to self-employed being in charge of their own working time, but it may also reflect that self-employed are building up inventories and capital stocks or have other means of extracting income from their firm (possibly also in the grey area between firm economics and personal economics). Finally, an institutional feature ("Virksomhedsordningen") allows self-employed to retain earnings in the firm.

[1]We compute the gross income including wage-income, capital income, labor market contributions (since 1994), taxable and non-taxable benefits. In order to obtain the disposable income we subtract the tax payments.

<< Figure 2 about here >>

The bunching at the tax brackets suggests that adding retained earnings (less of taxes) to the disposable income constitutes a better income measure for self-employed and we only use this income measure in the rest of the paper. As shown in Figure 3 we find that the unconditional mean and median incomes are larger in self-employment than in wage-employment in contrast to the US evidence in Hamilton (2000) and in contrast to when applying the narrow income measure.

<< Figure 3 about here >>

## 3. Econometric Specification

The organization of this section, can be summarized as follows: First, we consider the estimation of conditional earnings functions using Vella and Verbeek (1998, 1999) sample selection model for panel data. Hereafter, we construct income uncertainty and skewness measures. Finally, using measures for expected income, uncertainty and skewness, we model the occupational choice in a conditional logit model.

3.1. **Earnings Conditional on Occupational Choice.** For each person we separately predict the disposable income including retained earnings from being self-employed and being wage employed. The chosen income measure is disposable income including retained earnings. We use unemployment benefits for the group of unemployed. For each occupation we model earnings as a simple log-linear mincer earnings equation

$$\ln y^*_{nt} = x_{nt}\beta + \alpha_n + \varepsilon_{nt} \tag{3.1}$$

where $n$ indexes individuals $(n = 1, .., N)$ and $t$ indexes time $(t = 1, ...., T)$; $y^*_{nt}$ is annual disposable income plus retained earnings, $\beta$ is a vector of unknown coefficients to be estimated, $x_{nt}$ is a vector covariates, $\alpha_n$ represents unobserved heterogeneity and $\varepsilon_{nt}$ is a normally distributed disturbance.

Since we observe earnings for the chosen occupational status only, the conditional earnings functions will in general be estimated on a non-random selected sample. There are several arguments, why self-selection may be an issue in the present context. In the Roy (1951) model the individual ex-ante knows her sector-specific productivity, and will select herself into the sector, where she is most productive. Furthermore, if the incomes in the two sectors are highly correlated, the most productive persons will select the sector with the largest dispersion of sector specific abilities, while the least productive will select the sector with the smallest dispersion.

The Danish labor market is characterized by a compressed wage structure as a consequence of the generous unemployment benefit level and a high degree of organization on both employer and worker sides. As argued by Malchow-Møller, Markusen, and Skaksen (2006) such institutional arrangement may well imply that the most productive are not paid according to their marginal product and, therefore, the most able may select themselves into self-employment. On the other hand, the least productive may not have a sufficiently high productivity to earn the minimum wage in paid employment. Consequently, marginalization may also push the least productive into self-employment. Blume, Ejrnæs, Nielsen, and Würtz (2005) argue that this indeed is the case for non-western immigrants in the Danish labor market.

Yet another type of selection, ex-post self-selection, arises in leaning models such as Jovanovic (1979) and Jovanovic (1984), where persons have no ex-ante knowledge of their productivity, but consecutively observe output realizations. Persons experiencing poor output realizations will quit and search for a new match.

To control for the selection problem we use the Vella and Verbeek (1998, 1999) dynamic panel data application of Heckman's two-step sample selection model. The selection is modelled as a dynamic random effects probit, which allows us to separate two sources of persistence in the occupational choice: Persistence as a result of unobserved heterogeneity and (true) state dependence. Since we do not observe the first occupational choice, we cannot assume that the initial observation of the occupational is truly exogenous. We use the Wooldridge (2005) way of handling the initial conditions problem and, thus, allow the unobserved heterogeneity to be correlated with the initial dependent variable.

We will now briefly explain the model.[2] We consider a model consisting of two equations, where the parameters of equation (3.1) are of primary interest, while the selection equation

[2]For a detailed treatment of the model see Vella and Verbeek (1999).

below is a reduced form equation for the occupational choice. The selection part of the model can be summarized as

$$
\begin{aligned}
(3.2)\qquad d_{nt}^{*} &= x_{nt}\gamma + \phi d_{nt-1} + \mu_n + \eta_{nt} \\
(3.3)\qquad d_{nt} &= \mathbf{1}\left(d_{nt}^{*} > 0\right) \\
(3.4)\qquad \ln y_{nt} &= \ln y_{nt}^{*} \text{ if } d_{nt} = \mathbf{1} \\
&= 0 \text{ (unobserved) otherwise}
\end{aligned}
$$

where $y_{nt}^{*}$ and $d_{nt}^{*}$ are latent endogenous variables with observed counterparts $y_{nt}$ and $d_{nt}$.

The equation of interest is assumed to have the usual error component structure, where $\alpha_n \sim iN\left(0, \sigma_\alpha^2\right)$ and $\varepsilon_{nt} \sim iN\left(0, \sigma_\varepsilon^2\right)$. For the selection equation we allow for unobserved heterogeneity through random individual effects, such that the selection equation has the following two-component error structure $\mu_n \sim iN\left(0, \sigma_\mu^2\right)$ and $\eta_{nt} \sim iN\left(0, \sigma_\eta^2\right)$. We allow for correlation between the individual effects as well as correlation between the idiosyncratic disturbances, that is $\text{cov}(\alpha_n, \mu_n) = \sigma_{\alpha\mu} \neq 0$ and $\text{cov}(\varepsilon_{nt}, \eta_{nt}) = \sigma_{\varepsilon\eta} \neq 0$. Finally, denote $\xi_{nt} = \alpha_n + \varepsilon_{nt}$ , $\upsilon_{nt} = \mu_n + \eta_{nt}$ , $\mathbf{x}_n = \left[x_{n1}, ..., x_{nT}\right]'$ and let $\boldsymbol{\upsilon}_n$ be a $T$ vector of $\upsilon_{nt}$.

Assume now

$$
\begin{aligned}
(3.5)\qquad & \boldsymbol{\upsilon}_n | x_n \sim iN\left(0, \sigma_\mu^2 \mathbf{ii}' + \sigma_\eta^2 \mathbf{I}\right) \\
(3.6)\qquad & E\left[\xi_{nt} | x_n, \upsilon_n\right] = \tau_1 \upsilon_{nt} + \tau_2 \bar{\upsilon}_n
\end{aligned}
$$

where $\bar{\upsilon}_n = T^{-1} \sum_{t=1}^{T} \upsilon_{nt}$ and where $\tau_1 = \sigma_{\varepsilon\eta}/\sigma_\varepsilon^2$ and $\tau_2 = T\left(\sigma_{\alpha\mu} - \sigma_{\varepsilon\eta}\sigma_\mu^2/\sigma_\varepsilon^2\right) / \left(\sigma_\eta^2 + T\sigma_\mu^2\right)$ are constants to be estimated and $\mathbf{i}$ is a column of ones. Note that equation (3.6) imposes strict exogeneity of $x_{nt}$, such that errors are assumed to be independent of future and lagged values of $x_{nt}$. To estimate the conditional mean for the dependent variable in the equation of interest, we condition on the chosen occupation

$$
E\left[\ln y_{nt} | x_n, d_{n0}, d_n\right] = x_{nt}\beta + E\left[\xi_{nt} | x_n, d_{n0}, d_n\right]
$$

where $E\left[\xi_{nt} | x_n, d_{n0}, d_n\right]$ is the selection bias induced by correlation between the errors in the two equations.

Under these assumptions, it can be shown that the conditional mean of the error-term from the selection equation, $E\left[\upsilon_{nt}|x_n, d_{n0}, d_n\right]$ can be estimated by the following expression

$$\tilde{\upsilon}_{nt} = \frac{1}{\int f\left(d_n|x_n, \mu_n\right) f\left(\mu_n\right) d\mu_n} \int \left(\mu_n + E\left[\eta_{nt}|x_n, \mu_n\right]\right) f\left(d_n|x_n, \mu_n\right) f\left(\mu_n\right) d\mu_n \tag{3.7}$$

This expression can be approximated by quadrature methods or simulation. Once we have estimated the reduced form parameters for the selection equation, we can easily simulate the conditional error $\tilde{\upsilon}_{nt}$.[3]

After computing $\tilde{\upsilon}_{nt}$ and the individual specific means $\bar{\upsilon}_n = \frac{1}{T_n}\sum_t^{T_n} \tilde{\upsilon}_{nt}$ we can estimate the following equation by the simple linear random effects model

$$\ln y_{nt} = x_{nt}\beta + \tilde{\upsilon}_{nt}\theta_1 + \bar{\upsilon}_n\theta_2 + \alpha_n + \varepsilon_{nt}$$

3.2. **Uncertainty and Skewness Measures.** For each category in our disaggregated education breakdown shown in Table A.1 we estimate the occupational-specific measures of variance and skewness of the income processes. This is done separately for men and women.

We divide the uncertainty into a permanent part relating to the variance of the individual time-constant $\alpha_n$ and into a transitory uncertainty relating to the time-varying error-term.[4] Among the covariates in $x_{nt}$ we have included 28 educational dummies.[5] We define $a_n = \exp\left(\alpha_n\right)$ and $e_{nt} = \exp\left(\varepsilon_{nt}\right)$ and compute the variance $R$ and the skewness $K$ for each education type $l$ by

$$R_l^\alpha = \tfrac{1}{N_l}\sum_{n=1}^{N_l}\left(a_{nl} - \bar{a}_l\right)^2 \qquad K_l^\alpha = \tfrac{1}{N_l}\sum_{n=1}^{N_l}\left(a_{nl} - \bar{a}_l\right)^3$$

$$R_l^\varepsilon = \tfrac{1}{T}\tfrac{1}{N_l}\sum_{n=1}^{N_j}\left(e_{nlt} - \bar{e}_l\right)^2 \qquad K_l^\varepsilon = \tfrac{1}{T}\tfrac{1}{N_l}\sum_{n=1}^{N_l}\left(e_{nlt} - \bar{e}_l\right)^3$$

By averaging the residuals only on education groups, we effectively assume that the income uncertainty does not depend on for example experience, which is obviously an

---

[3]The procedure is summarized in algorithm 1 in the appendix.

[4]Recently, Diaz-Serrano, Hartog, and Nielsen (2003) have used a similar approach in the context of educational choice.

[5]In the IDA database there are 1,750 different educations, but in order to secure representativity we operate with 28 education groups only (see Table A.1 in the Appendix). We have aimed at securing representativity by not making a too disaggregated educational break-down, but on the other hand aimed at selecting as homogeneous groups as possible.

approximation. Averaging the incomes on other variables as well is not feasible with the detailed education break-down used.

For an unemployed there is no or very little uncertainty regarding income. Consequently, we set the variance and skewness equal to zero.

3.3. **A Model of Occupational Choice.** The behavioral framework underlying the occupational choice model is simple: We assume that individuals each period associate each occupation with a continuous random utility function, $U_{nit}$ , where each occupation is indexed by $i \in [se, we, ue]$. Each period individuals choose between self-employment (se), wage-employment (we) and unemployment (ue) to maximize the $U_{nit}$.[6] Random utility is assumed to be a linear function of occupational specific earnings, and the variance and skewness of permanent and transitory income shocks. Hence, the random utility function can be written as

$$U_{nit} = \mathbf{x}_{nit}\boldsymbol{\beta} + \delta_i + \epsilon_{nit} \quad \text{with} \quad n = 1, ..., N \quad \text{and } t = 1, .., T$$

where $\delta_i$ is a choice-specific constant, $\mathbf{x}_{nit} = \left[\hat{Y}_{nit}, R_l^{\alpha}, K_l^{\alpha}, R_l^{\varepsilon}, K_l^{\varepsilon}\right]$ denotes the set of attributes associated with each occupation, $\boldsymbol{\beta}$ is a vector of coefficients related to the the choice specific attributes $\mathbf{x}_{int}$. The error component $\epsilon_{nit}$ is assumed to be individual-, choice,- and time specific and distributed according to a Type I extreme value distribution. With this distributional assumption, we end up with McFadden's well known Conditional Logit model for discrete choices.

## 4. Results

4.1. **Self-selection and Earnings Differentials.** In this section, we investigate the extent to which earnings differentials can be explained by individuals self-selecting themselves into the different occupations. To account for the potentially important selection problems, we estimate the model sample selection model of Vella and Verbeek (1998,1999). First, we estimate the selection equation given by equation (3.2) and equation (3.3) by a

[6]Even though each individual maximizes utility each period by choosing occupations this need not be equivalent to maximization of life-time utility given by a discounted sum of period utility. However, this simplification is needed to make the model operational.

dynamic random effects probit. Hereafter, we estimate the parameters in the conditional earnings function in equation (3.1).

Since the choice of labor market state differs considerably between the genders, the sample correction and the prediction of incomes are done separately for men and women. Additionally, the existence of wage differentials between the genders suggests that it would be appropriate to run the wage equations separately.

The results from the selection equations given in table A.2 suggest that the impact of the lagged dependent variable is positive and highly significant, indicating the presence of substantial state dependence. State dependence can be a result of cost of and uncertainty of labor market transitions and is likely to be amplified for transitions into self-employment in the presence of start-up costs. In an intertemporal model of occupational choice Schjerning (2005) shows that the combination of irreversible start-up costs and income uncertainty introduces an option value of being self-employed. To avoid potential start-up costs associated with later re-entry, the self-employed is willing to wait until good times occur rather than temporarily leaving self-employment. This introduces a value of waiting and consequently we will see later entry and later exit.

The magnitude of state dependence for the self-employed is substansial: Being self-employed in the previous period increases the probability of being self-employed in the current period from 1.2 to 41.5 per cent for females and from 4.0 per cent to 48.7 per cent for men. As a comparison, the marginal effect from previous wage-employment is 19.9 per cent for females and 24.0 per cent for men.

The results from the selection equation suggest that, in general, the probability of being self-employed varies much between the educational categories both with respect to length and type of education. Although the picture is quite mixed, it seems to be the case that unskilled and some groups of highest education are the most likely to become self-employed. The latter is due to the fact that the self-employed include professionals such as practitioner doctors, dentists, lawyers and accountants.

The estimated earnings equations are given in table A.3. As dependent variable we use the disposable income including retained earnings. We allow for unobserved heterogeneity in the form of random effects, and we control for the usual socio-demographic variables. We find positive coefficients on marriage for men, while they are negative for women. The origin variables have the expected signs and magnitudes, i.e., non-western immigrants earn considerably less than western immigrants, second generation immigrants and natives. It is striking that non-western immigrants are more likely to become self-employed even though they should expect a much lower income in self-employment compared to wage-employment.

We find the usual hump-shaped effect of age, which obviously captures labor market experience. We have included dummies for each education from our detailed educational break-down shown in the appendix. The general picture is as expected that the longer education, the higher disposable income. As one would presume, the returns to education differ remarkably between the educations. For example, the returns to humanities are lower compared to social sciences at each length of education reflecting the relatively higher unemployment rate that may lead to accepting jobs below the educational level.

If education is a signal, so that employers use education to screen potential workers, we would expect lower returns to education in self-employment. There does not seem to be much evidence for the signalling hypothesis.

Since we do not wish to rely on the non-linearity of the selection equation to identify the selection effects in the income equations we need to exclude at least one variable. We use the lagged dependent variable, household wealth, dummies for children in the household and a dummy for the spouse being self-employed.

The inclusion of the correction terms account for the selection bias induced by the correlation between unobservables in the selection model and earnings equations. The coefficients to the correction terms $\upsilon_{nt}$ and $\bar{\upsilon}_n$ are statistically significant in all four regressions. In the case of men, the coefficient on both correction terms are negative, implying that the marginalization on average dominates. Taken literally, we have that those in wage-employment will tend to earn more in self-employment than those already self-employed.

In contrast to this, the coefficient on the individual specific correction term, $\bar{\upsilon}_n$ is positive in the self-employment earnings equation for women implying that those in wage-employment have a lower self-employment potential than the currently self-employed. Since the income is measured on a yearly basis a possible explanation for the positive selection into self-employment relates to differences in the hours of work between wage-employed and self-employed. About 20 per cent of female wage-employed work part-time and if this fraction is larger than the corresponding for women in self-selection a positive selection into self-employment will, on average, emerge. In recent work by Carrasco and Ejrnæs (2003) it is in fact argued that the relative low share of female self-employed in Denmark can be explained by the relative high level of flexibility in the Danish labor market providing the possibility to work part time in paid employment. Similar arguments apply to women planning to have children, as the opportunities for paid maternity leave are better in wage-employment. Another explanation might be glass-ceiling effects in wage-employment, see e.g. Albrecht, Björklund, and Vroman (2003) for Swedish evidence.

4.2. **The Occupational Choice.** The occupational choice model is estimated for a several different subsamples. The results from these estimations are shown in Table 2. Each row in the Table corresponds to the results for a different subsample. The figures in the Table show the effects of the mean, variance (uncertainty) and skewness (i.e., in this case the chance of very high incomes) of predicted earnings conditional on the occupational choice. Note that in the estimations, variance and skewness of both transitory and permanent shocks are included in the model. For expositional purposes, however, we only report variance and skewness of the permanent income component in the Table.[7]

<< Table 2 about here >>

The coefficients to mean earnings gives the marginal utility of expected income, while the (negative of the) coefficient to the variance can be interpreted as the marginal (dis-) utility of income uncertainty. To give an example, a positive coefficient to expected income is associated with individuals consistently choosing occupations with higher levels

[7]Since the earnings equations were estimated with age variables and time-dummies there is no aggregate time variation left in the error-terms, but still individual specific variation occurs. Alvarez, Browning, and Ejrnæs (2002) find that Danish income processes are particularly heterogenous. Steep income-experience profiles imply a large variance, but when controlling for the income level, we should due to income smoothing expect that a flat income-tenure profile is preferred. However, a steeper wage-experience profile may indicate greater possibilities such as promotion for wage-employed and business expansion for self-employed, which may explain the positice coefficient to the variance of the temporary income shocks. The skewness of the time-varying part does not seem to play any role.

of expected earnings, while a negative coefficient to variance emerges when individuals choose occupations with little income uncertainty.

To make results comparable across different subsamples, we compute the marginal rate of substitution (MRS) between the variance and mean earnings and between the skewness and mean earnings. The MRS can be interpreted as the rate at which you are willing to trade off more uncertainty for higher income.[8] These results are shown in the right part of the Table.

Considering the full sample (the first row in the Table), the results point to a large role for monetary aspects when choosing occupation. As expected, people's choice of occupation is positively affected by expected (mean) earnings and negatively by a higher variance of the income. Thus, on average, people appear risk averse. These findings are found to be robust to various sample decompositions.

Turning to the differences between the genders, we find that men put more emphasis on the earnings level, while women appear more risk averse. This is reflected in the much lower value of the MRS estimate for women. This could be one of the main reasons why fewer women choose to become self-employed.

Women seem to be behaving in a less risk averse manner when household wealth exceeds DKK 500,000. This is perfectly consistent with standard models of intertemporal behavior that find that the degree of effective risk aversion is decreasing in wealth; see, e.g., Deaton (1991), Carroll (1997), and Schjerning (2005).

The finding that married women appear less risk averse than other women is also fully consistent with models from the literature on family economics that point to risk sharing as being a potentially important economic gain from marriage, see e.g. Hess (2004).

A similar variation in men's attitudes towards risk is not found. An interesting finding, however, is that the MRS between income uncertainty and expected earnings is virtually zero compared to females. This confirms the evidence from Danish questionnaires, referred to above, which pointed to men putting much more emphasis on monetary gains (expected income) than women.

[8]This normalization is important since estimates from two different subsamples are not directly comparable due to differences in the variances of the unobserved factors.

Finally, a positive coefficient to skewness is interpreted as being consistent with evidence of overconfidence. If people systematically prefer occupations with a high degree of skewness (a chance of very high incomes) it may be due unrealistic, strong beliefs in their own ability.

For the full sample, a negative coefficient to skewness is found. Hence, on average, there is no evidence of overconfidence. The more detailed results with respect to this behavioral hypothesis are mixed and inconclusive. If anything, men behave somewhat more overconfidently than females. This result match those found from experimental studies.

It is striking that we in the model for immigrants only can explain 17 per cent compared to 50 per cent in the other models. Moreover, the coefficient to income is much lower than in the other conditional logit models. Hence, other important (unobserved) factors, such as lack of opportunities in the ordinary labor market and non-pecuniary benefits may be much more relevant in explaining their occupational choice. Hence, the low explanatory power, and the lower coefficient to income points to self-employment being the last resort due to marginalization in wage-employment. We also find that non-western immigrants appear less risk averse. This may be due to marginalization forcing immigrants to accept insecure and low paid occupations, but it can also be a consequence of cultural differences in the attitudes towards self-employment.

## 5. Conclusions

This paper uses high quality Danish longitudinal register data, to investigate the relationship between self-employment choice, expected earnings and income uncertainty. We proceed in the following steps: Firstly, we estimate of conditional earnings functions using the sample selection model of Vella and Verbeek (1998, 1999). Secondly, using measures for expected income, uncertainty and skewness, we model the occupational choice in a conditional logit model.

Comparing earnings distributions based on different income measures, we find that i) the dispersion of incomes is in general much larger for the self-employed and ii) Danish self-employed earn more than wage-employed when retained earnings are included in the income measure. Contrary to wage-workers, self-employed (taxable) personal income bunch at kink points in the tax system since self-employed (unlike wage workers) has the possibility to retain earnings and thereby transfer income across years. The progressive Danish income tax system provides strong incentives to make such transfers.

Several experimental studies have found that while men are more competitive, women are more risk averse. In the context of occupational choice, we find that men put more emphasis on the income level, while women seem to be more risk-averse. This result is found to be robust to various sample decompositions.

Linking the behavioral results from the experimental literature with income distributions in self-employment and wage-employment may explain why fewer women become self-employed. We find that part of the gender gap can be explained by gender differences in the trade-offs between income level and the variance of incomes. However, we find no effect of skewness of incomes.

Non-western immigrants are overrepresented in self-employment. The occupational choice model performs considerably worse for this group and we find smaller effects of income level and variance. Furthermore, the sample selection model shows that non-western are more likely to become self-employed even though they should expect a much lower income in self-employment than native Danes. This suggests that non-western immigrants are marginalized into self-employment.

## 6. Appendix

**Algorithm 1.** *Estimation of conditional error-term from the selection equation,* $E\left[\upsilon_{nt}|x_{nt},d_{n0},d_{nt}\right]$

(1) *For a given set of parameter values* $\theta_1 = (\gamma,\phi,\sigma_\mu)$ *take a draw from* $\mu_n^r$ *from* $f(\mu_n|\sigma_\mu) = N(0,\sigma_\mu)$ *and calculate the likelihood for individual i conditional on the draw*

$$f\left(d_n,d_{i0}|x_{n,}\mu_n^r\right) = \prod_{t=1}^{T_n} f\left(d_{nt}|x_{nt,}\mu_n^r\right) f\left(d_{n0}|x_{nt,}\mu_n^r\right)$$

*where* $f\left(d_{nt}|x_{nt,}\mu_n^r\right) = \Phi_{nt}d_{nt}+(1-\Phi_{nt})(1-d_{nt})$ *and where* $\Phi_{nt} \equiv \Phi\left(x_{nt}\gamma+\phi d_{nt-1}+\mu_n^r\right)$

(2) *Repeat many times and average the results to obtain the Simulated Log Likelihood function (SLL)*

$$SLL = \ln\frac{1}{R}\sum_r^R f\left(d_n,d_{i0}|x_{n,}\mu_n^r\right)$$

(3) *Choose* $\theta_1^{MSL}$ *so that SLL is maximized*

(4) *Given the MSL estimates from the fist stage regression* $\theta_1^{MSL}$, *we can easily simulate* $\hat{\upsilon}_{nt}$. *Take R draws from* $f\left(\mu_n|\sigma_\mu^{MSL}\right)$ *and calculate the simulated counterpart of* $\hat{\upsilon}_{nt}$

$$\tilde{\upsilon}_{nt} = \frac{1}{\frac{1}{R}\sum_r^R f\left(d_n,d_{i0}|x_{n,}\mu_n^r\right)}\frac{1}{R}\sum_r^R\left(\mu_n^r+E\left[\eta_{nt}|x_n,\mu_n^r\right]\right) f\left(d_n,d_{i0}|x_{n,}\mu_n^r\right)$$

*where* $E\left[\eta_{nt}|x_n,\mu_n^r\right] = \frac{d_{nt}\phi_{nt}}{\Phi_{nt}}\text{-}\frac{(1-d_{nt})\phi_{nt}}{1-\Phi_{nt}}$ *is the cross-sectional generalized residual for the probit model and where* $\phi_{nt} \equiv \phi\left(x_{nt}\gamma+\phi d_{nt-1}+\mu_n^r\right)$

To improve coverage of the integrals and reduce simulation noise, we use Halton Draws.[9]

.

[9]Halton draws provides a superior coverage as it induces negative correlation across individuals. In the context of discrete choice models, Bhat (2001) found in a Mixed Logit Model, that 100 Halton draws provided more precise results than 1000 standard pseudo random draw. Train (2003) provide a comprehensive and excellent treatment of several variance reduction techniques.

## References


Albrecht, J., A. Björklund, and S. Vroman (2003): "Is There a Glass Celing in Sweden?," *Journal of Labor Economics*, 21(1), 145–177.

Alvarez, J., M. Browning, and M. Ejrnæs (2002): "Modelling Income Processes with Lots of Heterogeneity," CAM Working Paper 2002-01.

Bhat, C. (2001): "Quasi-Random Maximum Simulated Likelihood Estimation of the Multinomial Mixed Logit Model," *Transportation Research*, B 35, 677–693.

Blume, K., M. Ejrnæs, H. S. Nielsen, and A. Würtz (2005): "Labor Market Transitions of Immigrants with Emphasis on Marginalization and Self-Employment," working paper.

Byrnes, J. P., D. C. Miller, and W. D. Schafer (1999): "Gender Differences in Risk-Taking: A Meta-Analysis," *Psychological Bulletin*, 125, 367–383.

Carrasco, R., and M. Ejrnæs (2003): "Self-Employment in Denmark and Spain: Institution, Economic Conditions and Gender Differences," CAM Working Paper 2003-06.

Carroll, C. D. (1997): "Buffer Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quaterly Journal of Economics*, 112(1), 1–56.

Coelho, M., and D. de Meza (2006): "Self Deception, Self Selection, Self Destruction: An Experimental Investigation of Adverse Selection," working paper.

Croson, R., and U. Gneezy (2004): "Gender Differences in Preferences," working paper.

Deaton, A. (1991): "Savings and Liquidity Constraints," *Econometrica*, 59, 1121–1248.

Diaz-Serrano, L., J. Hartog, and H. S. Nielsen (2003): "Compensating Wage Differentials for Schooling Risk in Denmark," IZA Discussion Paper no. 963.

Elston, J. A., G. W. Harrison, and E. E. Rutström (2006): "Characterizing the Entrepreneur Using Field Experiments," working paper.

Hamilton, B. H. (2000): "Does Entrepreneurhip Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, 108, 604–631.

Hess, G. D. (2004): "Marriage and Consumption Insurance: What's Love Got to Do with It?," *Journal of Political Economy*, 112, 290–318.

Jovanovic, B. (1979): "Job Matching and theTheory of Turnover," *Journal of Political Economy*, 87, 972–990.

——— (1984): "Matching, Turnover and Unemployment, Journal of Political Economy," *Journal of Political Economy*, 92, 108–122.

Malchow-Møller, N., J. R. Markusen, and J. R. Skaksen (2006): "Labour Market Institutions, Learning, and Self-Employment," CEBR discussion paper 2006-24.

Niederle, M., and L. Vesterlund (2006): "Do Women Shy Away from Competition? Do Men Compete Too Much?," working paper.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Schjerning, B. (2005): "A Dynamic Model for Occupational Choice, Savings and Formation of Entrepreneurial Human Capital under Liquidity Constraints and Income Uncertainty," working paper.

StatisticsDenmark (1999): "Generel Erhvervsstatistk," Statistiske Efterretninger 1999:19.

Svenson, O. (1980): "Are We All Less Risky and More Skillful Than our Fellow Drivers?," *Acta Psychologica*, 47, 143–148.

Train, K. (2003): *Discrete Choice Models with Simulation*. Cambridge University Press, Cambridge.

Vella, F., and M. Verbeek (1998): "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men," *Journal of Applied Econometrics*, 13(2), 163–183.

Vella, F., and M. Verbeek (1999): "Two-Step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias," *Journal of Econometrics*, 90, 239–263.

Wooldridge, J. M. (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Effects," *Journal of Applied Econometrics*, 20(1), 39–54.

*Department of Economics, University of Copenhagen*
*E-mail address*: `christian.daniel.le.maire@econ.ku.dk`
*URL*: `http://www.lemaire.dk`

*Centre for Economic and Business Research (CEBR)*
*E-mail address*: `bsc.cebr@cbs.dk`
*URL*: `http://www.cebr.dk/bsc`

**FIGURE 1: DISTRIBUTION OF DISPOSABLE INCOME IN 1996**

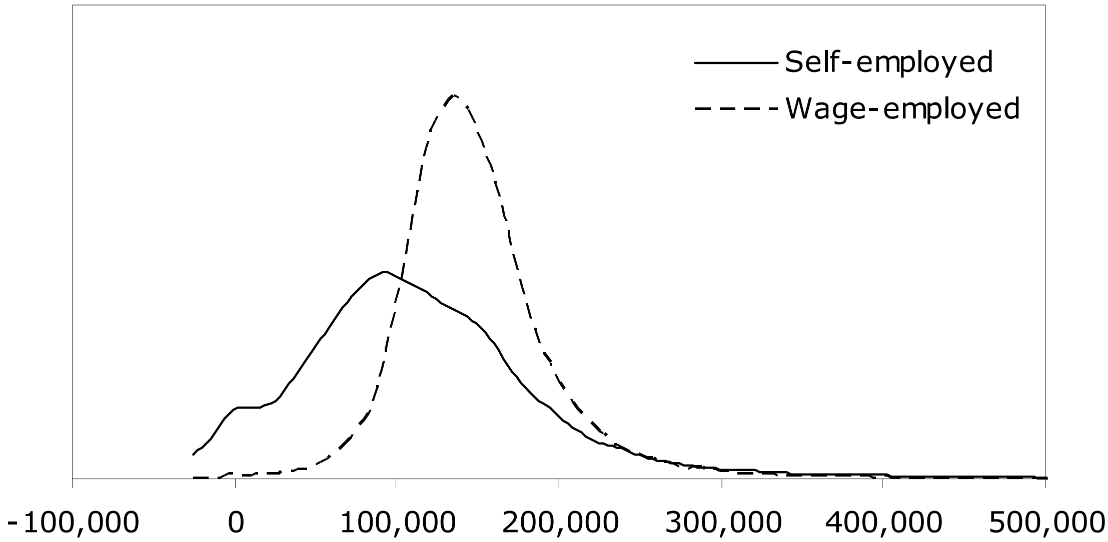


**TABLE 1: INCOME DISTRIBUTIONS IN 1996, SELECTED PERCENTILES**

| Percentile | Personal Income | | Disposable Income | |
|---|---|---|---|---|
| | Self-empl. | Wage-empl. | Self-empl. | Wage-empl. |
| 5 | 0 | 111,055 | -12,529 | 85,848 |
| 10 | 7,745 | 135,380 | 12,687 | 99,698 |
| 25 | 75,980 | 167,222 | 61,119 | 119,634 |
| 40 | 120,678 | 190,022 | 88,076 | 133,592 |
| 50 | 148,590 | 205,274 | 104,021 | 142,430 |
| 60 | 183,967 | 221,686 | 121,745 | 151,914 |
| 75 | 242,834 | 253,018 | 152,229 | 169,066 |
| 90 | 345,187 | 325,042 | 202,619 | 202,231 |
| 95 | 458,649 | 387,028 | 248,060 | 229,819 |

**FIGURE 2: DISTRIBUTION OF PERSONAL INCOME IN 1996**

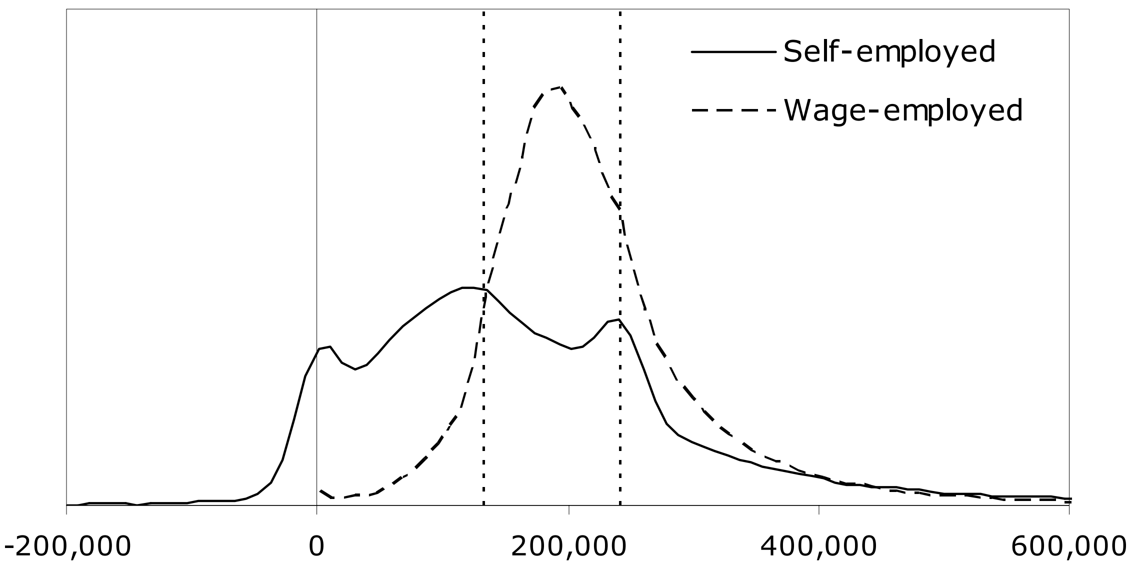

**FIGURE 3: DISTRIBUTION OF DISPOSABLE INCOME PLUS RETAINED EARNINGS**

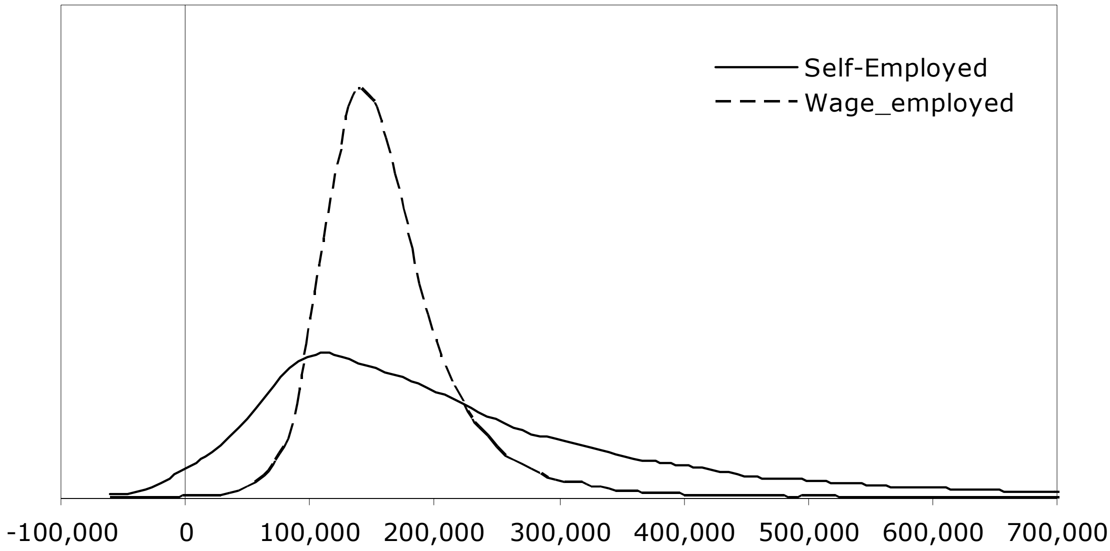


**TABLE 2: CHOICE OF LABOR MARKET STATUS**
**(CONDITIONAL LOGIT MODEL)**

| | | | Distribution of Permanent Chock's | | | | MRS | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Subsample | Mean Earnings | | Variance | | Skewness | | Variance/ Mean | Skewnes/ Mean | Sample size | $R^2$ |
| All | 1.58 | (0.007)** | -0.15 | (0.004)** | -0.02 | (0.001)** | -0.10 | -0.010 | 7,274,082 | 51 |
| Non-western immigrants | 1.08 | (0.044)** | 0.05 | (0.025) | -0.02 | (0.005)** | 0.04 | -0.016 | 119,412 | 23 |
| Women | 0.40 | (0.017)** | -0.22 | (0.011)** | -0.04 | (0.001)** | -0.56 | -0.091 | 3,391,905 | 59 |
| Non-western immigrants | 0.81 | (0.095)** | -0.34 | (0.070)** | 0.02 | (0.009)* | -0.42 | 0.022 | 45,564 | 29 |
| Married | 1.00 | (0.023)** | -0.27 | (0.014)** | -0.03 | (0.002)** | -0.27 | -0.032 | 2,514,129 | 61 |
| HH. Wealth(t-1)>500.000 | 0.85 | (0.048)** | -0.14 | (0.022)** | -0.01 | (0.003)** | -0.16 | -0.013 | 493,101 | 64 |
| Father self-employed | 1.29 | (0.148)** | 0.20 | (0.030)** | -0.06 | (0.009)** | 0.16 | -0.043 | 64,728 | 64 |
| age<40 | 0.40 | (0.026)** | -0.11 | (0.013)** | -0.05 | (0.002)** | -0.27 | -0.118 | 1,353,195 | 60 |
| age>45 | 0.58 | (0.030)** | -0.37 | (0.025)** | -0.02 | (0.002)** | -0.63 | -0.038 | 1,148,580 | 57 |
| Men | 2.81 | (0.016)** | -0.10 | (0.005)** | 0.00 | (0.001) | -0.04 | -0.001 | 3,882,177 | 44 |
| Non-western immigrants | 1.81 | (0.088)** | 0.14 | (0.027)** | -0.03 | (0.006)** | 0.08 | -0.018 | 73,848 | 20 |
| Married | 2.39 | (0.020)** | -0.15 | (0.006)** | 0.01 | (0.001)** | -0.06 | 0.003 | 2,761,398 | 48 |
| HH. Wealth(t-1)>500.000 | 1.42 | (0.044)** | -0.18 | (0.012)** | 0.06 | (0.002)** | -0.13 | 0.045 | 474,390 | 43 |
| Father self-employed | 3.77 | (0.110)** | -0.10 | (0.023)** | -0.02 | (0.007)** | -0.03 | -0.006 | 91,098 | 40 |
| age<40 | 3.33 | (0.027)** | -0.07 | (0.007)** | -0.02 | (0.002)** | -0.02 | -0.006 | 1,538,208 | 50 |
| age>45 | 2.53 | (0.027)** | -0.10 | (0.009)** | 0.01 | (0.002)** | -0.04 | 0.004 | 1,348,191 | 39 |

Notes: Standard errors are in parentheses. * significant at 5%; ** significant at 1%. Other controls: Occupational specific constants and measures of the temporary components of estimated chocks (skewness and variance)

## TABLE A.1: MEAN, VARIANCE AND SKEWNESS (EDUCATIONAL BREAKDOWN)

| | | # observations | | Mean disposable income | | Variance | | | | Skewness | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Transitory effect | | Permanent effect | | Transitory effect | | Permanent effect | |
| | | se | we | se | we | se | we | se | we | se | we | se | we |
| Missing education | | 3,867 | 29,167 | 209,691 | 163,169 | 35.633 | 0.049 | 2.196 | 0.224 | 47.9 | 11.8 | 6.1 | 3.9 |
| Primary School | *Basic school* | 80,596 | 657,151 | 217,842 | 133,278 | 35.476 | 0.037 | 2.296 | 0.099 | 236.9 | 36.9 | 25.8 | 7.5 |
| Secondary school | *General* | 4,509 | 43,962 | 267,294 | 166,541 | 0.660 | 0.039 | 3.216 | 0.223 | 25.5 | 2.1 | 6.9 | 3.6 |
| | *Commercial and technical* | 1,362 | 12,906 | 353,176 | 218,601 | 1.062 | 0.100 | 6.970 | 0.806 | 11.9 | 20.8 | 6.9 | 13.1 |
| Vocational training | *Shop assistents* | 24,343 | 345,545 | 232,634 | 151,299 | 2.342 | 0.034 | 0.974 | 0.122 | 104.9 | 9.8 | 5.3 | 6.2 |
| | *Building and construction* | 18,769 | 113,541 | 225,676 | 170,977 | 0.571 | 0.024 | 0.554 | 0.072 | 28.6 | 5.1 | 7.5 | 4.4 |
| | *Metal* | 17,887 | 147,347 | 240,837 | 177,130 | 0.219 | 0.028 | 0.784 | 0.075 | 11.2 | 19.9 | 5.7 | 4.5 |
| | *Graphic* | 2,061 | 18,304 | 277,758 | 202,616 | 3.609 | 0.029 | 2.278 | 0.161 | 22.1 | 4.5 | 7.2 | 8.3 |
| | *Technical* | 2,606 | 33,878 | 158,173 | 129,978 | 1.342 | 0.035 | 1.147 | 0.097 | 18.8 | 14.5 | 3.0 | 2.7 |
| | *Service and transport* | 10,725 | 23,617 | 124,194 | 131,480 | 1.225 | 0.051 | 0.478 | 0.095 | 42.1 | 22.1 | 1.7 | 1.5 |
| | *Food* | 20,383 | 49,809 | 314,211 | 170,244 | 1.155 | 0.032 | 0.764 | 0.091 | 83.8 | 5.1 | 5.3 | 4.7 |
| | *Health care* | 2,254 | 67,273 | 138,455 | 117,762 | 0.659 | 0.026 | 0.896 | 0.052 | 14.2 | 5.2 | 3.3 | 1.2 |
| Post secondary | *Humanities and social sciences* | 1,576 | 18,953 | 185,609 | 153,600 | 0.224 | 0.039 | 2.013 | 0.141 | 2.1 | 2.5 | 3.6 | 4.5 |
| | *Technical* | 3,847 | 32,146 | 246,378 | 181,377 | 0.202 | 0.040 | 0.594 | 0.112 | 3.7 | 15.2 | 3.0 | 6.7 |
| | *Agriculture* | 917 | 8,504 | 306,245 | 157,736 | 0.218 | 0.039 | 0.697 | 0.066 | 3.2 | 5.1 | 1.1 | 1.1 |
| | *Health care* | 206 | 10,378 | 114,543 | 129,392 | 1.163 | 0.024 | 0.812 | 0.051 | 4.5 | 2.0 | 0.3 | 0.9 |
| | *Police and defence* | 327 | 15,524 | 234,734 | 202,600 | 0.194 | 0.020 | 0.621 | 0.079 | 3.7 | 6.2 | 3.2 | 3.8 |
| Higher education short cycle | *Humanities* | 3,033 | 157,352 | 174,274 | 156,319 | 7.537 | 0.021 | 0.893 | 0.054 | 48.4 | 2.6 | 3.1 | 1.5 |
| | *Social sciences* | 1,780 | 22,332 | 491,580 | 251,898 | 0.251 | 0.073 | 1.292 | 0.270 | 3.2 | 15.7 | 2.5 | 3.8 |
| | *Technical* | 3,338 | 40,915 | 308,715 | 246,404 | 0.483 | 0.041 | 1.327 | 0.126 | 17.3 | 8.7 | 8.4 | 10.5 |
| | *Health care* | 1,962 | 65,014 | 168,011 | 137,621 | 0.095 | 0.026 | 0.659 | 0.057 | 2.0 | 2.7 | 4.8 | 4.7 |
| | *Food, agriculture and transport* | 730 | 15,661 | 276,414 | 216,992 | 0.883 | 0.032 | 2.391 | 0.073 | 19.7 | 2.4 | 5.1 | 1.7 |
| | *BA* | 469 | 5,205 | 336,057 | 251,710 | 0.828 | 0.087 | 1.234 | 0.320 | 8.4 | 11.9 | 1.5 | 2.4 |
| Higher education MA level | *Humanities* | 578 | 22,592 | 209,513 | 190,457 | 0.162 | 0.038 | 0.868 | 0.068 | 2.3 | 26.6 | 1.7 | 1.5 |
| | *Natural sciences* | 159 | 9,727 | 226,141 | 217,852 | 0.179 | 0.023 | 1.104 | 0.070 | 2.8 | 1.2 | 4.0 | 3.8 |
| | *Social sciences* | 3,589 | 25,269 | 476,302 | 251,683 | 0.321 | 0.047 | 0.778 | 0.190 | 19.3 | 6.9 | 3.5 | 5.4 |
| | *Technical* | 1,912 | 18,252 | 334,441 | 255,821 | 0.553 | 0.048 | 2.193 | 0.161 | 7.1 | 17.9 | 10.3 | 7.0 |
| | *Food* | 1,431 | 5,250 | 358,845 | 228,455 | 0.122 | 0.024 | 0.446 | 0.073 | 5.8 | 2.1 | 7.6 | 2.4 |
| | *Health care* | 7,069 | 18,412 | 441,567 | 260,069 | 0.074 | 0.038 | 0.199 | 0.085 | 2.4 | 13.1 | 3.0 | 1.3 |

## TABLE A.2: SELECTION EQUATIONS
## (RESULTS FROM A BINARY PROBIT WITH RANDOM EFFECTS)

| | | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Self-employment | | Wage-employment | | Self-employment | | Wage-employment | |
| | | Coefficient | Std. | Coefficient | Std. | Coefficient | Std. | Coefficient | Std. |
| Lagged dependent, y(t-1) | | 2.590 | 0.007 | 1.415 | 0.004 | 2.651 | 0.010 | 1.226 | 0.005 |
| Initial dependent, y(0) | | 1.962 | 0.018 | 1.918 | 0.010 | 1.641 | 0.023 | 1.540 | 0.011 |
| Age | | 0.705 | 0.071 | -0.157 | 0.044 | 0.731 | 0.096 | 0.974 | 0.047 |
| Age squared | | -0.827 | 0.082 | 0.067 | 0.050 | -0.822 | 0.110 | -1.231 | 0.053 |
| Wealth (in mio dkr, 1996 prices) | | 0.014 | 0.002 | 0.001 | 0.001 | 0.005 | 0.002 | 0.008 | 0.001 |
| No. of children aged 0-6 | | 0.049 | 0.006 | -0.019 | 0.004 | 0.049 | 0.010 | -0.098 | 0.006 |
| No. of children aged 7-17 | | 0.030 | 0.004 | -0.011 | 0.003 | 0.012 | 0.006 | 0.001 | 0.003 |
| Married | | -0.006 | 0.009 | 0.112 | 0.007 | -0.047 | 0.011 | 0.064 | 0.007 |
| Immigrant (western) | | 0.043 | 0.030 | -0.155 | 0.023 | 0.132 | 0.032 | -0.159 | 0.021 |
| Immigrant (non-western) | | 0.230 | 0.028 | -0.481 | 0.020 | 0.124 | 0.037 | -0.529 | 0.024 |
| Second generation immigrants | | 0.092 | 0.084 | -0.111 | 0.070 | 0.132 | 0.109 | 0.037 | 0.079 |
| Spouse self-employed | | 0.453 | 0.013 | 0.137 | 0.005 | 0.399 | 0.011 | 0.152 | 0.005 |
| Regional dummies | *Copenhagen* | 0.009 | 0.017 | -0.131 | 0.012 | 0.033 | 0.022 | -0.097 | 0.013 |
| | *Large city* | 0.012 | 0.016 | -0.096 | 0.012 | 0.022 | 0.020 | -0.161 | 0.012 |
| | *Rural* | 0.097 | 0.013 | -0.123 | 0.009 | 0.120 | 0.015 | -0.197 | 0.009 |
| Missing education | | 0.066 | 0.029 | -0.030 | 0.023 | 0.152 | 0.037 | -0.109 | 0.025 |
| Secondary school | *General* | 0.061 | 0.026 | 0.073 | 0.020 | 0.133 | 0.034 | 0.093 | 0.022 |
| | *Commercial and technical* | 0.070 | 0.043 | 0.153 | 0.038 | 0.002 | 0.070 | 0.190 | 0.042 |
| Vocational training | *Shop assistents* | -0.038 | 0.014 | 0.176 | 0.012 | -0.023 | 0.013 | 0.155 | 0.008 |
| | *Building and construction* | -0.025 | 0.014 | 0.067 | 0.011 | 0.180 | 0.084 | -0.011 | 0.067 |
| | *Metal* | -0.072 | 0.014 | 0.152 | 0.011 | 0.403 | 0.139 | 0.063 | 0.103 |
| | *Graphic* | -0.025 | 0.036 | 0.041 | 0.028 | 0.252 | 0.082 | -0.182 | 0.055 |
| | *Technical* | -0.010 | 0.046 | 0.087 | 0.033 | 0.071 | 0.029 | 0.000 | 0.019 |
| | *Service and transport* | 0.188 | 0.041 | -0.145 | 0.032 | 0.500 | 0.025 | -0.370 | 0.020 |
| | *Food* | 0.175 | 0.018 | -0.156 | 0.015 | 0.110 | 0.057 | 0.051 | 0.037 |
| | *Health care* | -0.143 | 0.086 | 0.418 | 0.066 | -0.099 | 0.022 | 0.351 | 0.013 |
| Post secondary | *Humanities and social sciences* | 0.171 | 0.069 | 0.007 | 0.059 | 0.176 | 0.036 | 0.063 | 0.023 |
| | *Technical* | -0.056 | 0.026 | 0.191 | 0.021 | 0.057 | 0.054 | 0.223 | 0.035 |
| | *Agriculture* | -0.050 | 0.060 | 0.126 | 0.047 | -0.172 | 0.084 | 0.439 | 0.054 |
| | *Health care* | -0.751 | 0.361 | 0.478 | 0.153 | -0.204 | 0.064 | 0.443 | 0.038 |
| | *Police and defence* | -0.425 | 0.053 | 0.763 | 0.039 | -0.422 | 0.322 | 0.532 | 0.122 |
| Higher education short cycle | *Humanities* | -0.365 | 0.026 | 0.587 | 0.018 | -0.221 | 0.019 | 0.552 | 0.012 |
| | *Social sciences* | 0.003 | 0.035 | 0.256 | 0.029 | -0.097 | 0.057 | 0.418 | 0.037 |
| | *Technical* | -0.048 | 0.024 | 0.302 | 0.019 | 0.139 | 0.118 | 0.073 | 0.078 |
| | *Health care* | 0.153 | 0.077 | 0.285 | 0.065 | -0.177 | 0.024 | 0.785 | 0.016 |
| | *Food, agriculture and transportat* | -0.258 | 0.043 | 0.478 | 0.033 | -0.095 | 0.114 | 0.277 | 0.073 |
| | *BA* | 0.043 | 0.067 | 0.170 | 0.054 | -0.238 | 0.189 | 0.324 | 0.094 |
| Higher education MA level | *Humanities* | -0.358 | 0.047 | 0.510 | 0.032 | -0.044 | 0.055 | 0.297 | 0.032 |
| | *Natural sciences* | -0.390 | 0.069 | 0.616 | 0.051 | -0.221 | 0.122 | 0.352 | 0.065 |
| | *Social sciences* | 0.198 | 0.029 | 0.073 | 0.023 | 0.195 | 0.051 | 0.218 | 0.034 |
| | *Technical* | 0.021 | 0.033 | 0.264 | 0.027 | 0.112 | 0.101 | 0.110 | 0.072 |
| | *Food* | 0.223 | 0.053 | 0.034 | 0.042 | 0.395 | 0.094 | -0.008 | 0.062 |
| | *Health care* | 0.587 | 0.031 | -0.333 | 0.027 | 0.563 | 0.039 | 0.048 | 0.029 |
| | Year dummies | Yes | | Yes | | Yes | | Yes | |
| | Constant | -4.327 | 0.153 | -0.661 | 0.098 | -4.627 | 0.208 | -2.555 | 0.104 |
| | $\sigma_\mu$ | 0.760 | 0.009 | 0.838 | 0.005 | 0.626 | 0.011 | 0.730 | 0.005 |
| Number of observations | | 1288888 | | 1288888 | | 1126960 | | 1126960 | |
| Number of individuals | | 136990 | | 136990 | | 122749 | | 122749 | |
| Log-likelihood | | -116985.808 | | -287641.544 | | -59297.355 | | -243222.453 | |

## TABLE A.3: EARNINGS EQUATIONS
## (CORRECTED FOR SAMPLE SELECTION BIAS AND UNOBSERVED HETEROGENEITY)

| | | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Self-employment | | Wage-employment | | Self-employment | | Wage-employment | |
| | | Coefficient | Std. | Coefficient | Std. | Coefficient | Std. | Coefficient | Std. |
| Age (divided by 10) | | 0.707 | 0.030 | 0.323 | 0.004 | 0.282 | 0.081 | 0.069 | 0.005 |
| Age squared (divided by 1000) | | -0.834 | 0.034 | -0.350 | 0.005 | -0.410 | 0.092 | -0.111 | 0.005 |
| Married | | 0.110 | 0.006 | 0.063 | 0.001 | -0.285 | 0.015 | -0.101 | 0.001 |
| Immigrant (western) | | -0.278 | 0.031 | -0.081 | 0.006 | -0.264 | 0.062 | -0.046 | 0.006 |
| Immigrant (non-western) | | -0.608 | 0.029 | -0.187 | 0.006 | -0.543 | 0.071 | -0.065 | 0.008 |
| Second generation immigrants | | 0.008 | 0.086 | -0.027 | 0.014 | -0.098 | 0.230 | 0.009 | 0.018 |
| Regional | *Copenhagen* | -0.239 | 0.015 | -0.085 | 0.002 | -0.141 | 0.034 | -0.044 | 0.002 |
| dummies | *Large city* | 0.019 | 0.016 | -0.039 | 0.002 | -0.033 | 0.036 | -0.062 | 0.002 |
| | *Rural* | 0.041 | 0.012 | 0.014 | 0.002 | -0.072 | 0.027 | -0.042 | 0.002 |
| Missing education | | -0.068 | 0.031 | 0.086 | 0.006 | -0.098 | 0.072 | 0.063 | 0.007 |
| Secondary | *General* | 0.079 | 0.029 | 0.017 | 0.005 | 0.093 | 0.062 | 0.009 | 0.005 |
| school | *Commercial and technical* | 0.258 | 0.047 | 0.238 | 0.009 | -0.056 | 0.131 | 0.114 | 0.010 |
| Vocationa | *Shop assistents* | 0.196 | 0.016 | 0.185 | 0.003 | 0.107 | 0.029 | 0.093 | 0.002 |
| training | *Building and construction* | 0.025 | 0.015 | 0.066 | 0.003 | -0.152 | 0.194 | 0.052 | 0.017 |
| | *Metal* | 0.071 | 0.015 | 0.087 | 0.003 | 0.175 | 0.288 | 0.101 | 0.027 |
| | *Graphic* | 0.148 | 0.041 | 0.223 | 0.007 | 0.276 | 0.172 | 0.171 | 0.017 |
| | *Technical* | 0.038 | 0.051 | 0.042 | 0.009 | -0.113 | 0.061 | 0.036 | 0.005 |
| | *Service and transport* | 0.002 | 0.036 | 0.084 | 0.009 | -0.013 | 0.038 | -0.028 | 0.007 |
| | *Food* | 0.319 | 0.016 | 0.085 | 0.004 | 0.182 | 0.099 | 0.051 | 0.008 |
| | *Health care* | 0.144 | 0.112 | 0.043 | 0.012 | 0.049 | 0.051 | 0.039 | 0.003 |
| Post | *Humanities and social sciences* | 0.086 | 0.072 | 0.209 | 0.013 | 0.219 | 0.067 | 0.189 | 0.006 |
| secondary | *Technical* | 0.139 | 0.030 | 0.178 | 0.005 | 0.098 | 0.121 | 0.073 | 0.009 |
| | *Agriculture* | 0.316 | 0.064 | 0.218 | 0.013 | 0.071 | 0.171 | 0.135 | 0.013 |
| | *Health care* | 0.141 | 0.452 | -0.021 | 0.027 | -0.044 | 0.149 | 0.113 | 0.009 |
| | *Police and defence* | 0.242 | 0.076 | 0.208 | 0.007 | 0.215 | 0.755 | 0.086 | 0.022 |
| Higher | *Humanities* | 0.133 | 0.040 | 0.189 | 0.004 | 0.122 | 0.052 | 0.224 | 0.003 |
| education | *Social sciences* | 0.660 | 0.041 | 0.426 | 0.006 | 0.443 | 0.151 | 0.276 | 0.008 |
| short cycl | *Technical* | 0.300 | 0.029 | 0.401 | 0.005 | 0.324 | 0.207 | 0.385 | 0.020 |
| | *Health care* | 0.279 | 0.097 | 0.180 | 0.014 | 0.468 | 0.058 | 0.219 | 0.004 |
| | *Food, agriculture and transport* | 0.181 | 0.061 | 0.341 | 0.008 | 0.299 | 0.255 | 0.191 | 0.017 |
| | *BA* | 0.127 | 0.068 | 0.314 | 0.011 | 0.484 | 0.313 | 0.163 | 0.015 |
| Higher | *Humanities* | 0.233 | 0.080 | 0.336 | 0.007 | 0.247 | 0.109 | 0.445 | 0.007 |
| education | *Natural sciences* | 0.354 | 0.114 | 0.387 | 0.009 | -0.066 | 0.378 | 0.476 | 0.015 |
| MA level | *Social sciences* | 0.773 | 0.034 | 0.486 | 0.006 | 0.719 | 0.109 | 0.476 | 0.007 |
| | *Technical* | 0.376 | 0.039 | 0.477 | 0.007 | 0.564 | 0.198 | 0.435 | 0.020 |
| | *Food* | 0.634 | 0.054 | 0.428 | 0.013 | 0.613 | 0.200 | 0.420 | 0.023 |
| | *Health care* | 0.959 | 0.028 | 0.663 | 0.008 | 1.257 | 0.069 | 0.627 | 0.010 |
| $\sigma_{\alpha\mu}$ | | -0.062 | 0.003 | -0.112 | 0.001 | -0.024 | 0.005 | -0.076 | 0.002 |
| $\sigma_{\varepsilon\eta}$ | | -0.012 | 0.004 | 0.010 | 0.001 | 0.092 | 0.008 | 0.013 | 0.001 |
| Constant | | 15.109 | 0.068 | 15.783 | 0.010 | 15.779 | 0.180 | 16.323 | 0.011 |
| | *Summary Statistics* | | | | | | | | |
| | Number of observations | 168782 | | 1037089 | | 47496 | | 994058 | |
| | Number of individuals | 27544 | | 122754 | | 10456 | | 117511 | |
| | $\sigma_{\alpha}$ | 0.6656 | | 0.2929 | | 0.9831 | | 0.3092 | |
| | $\sigma_{\varepsilon}$ | 0.4557 | | 0.1737 | | 0.5914 | | 0.1869 | |
| | Fraction of variance due to individual specific error | 0.68 | | 0.74 | | 0.73 | | 0.73 | |
| | R-squared | 0.12 | | 0.20 | | 0.11 | | 0.23 | |