

Dagsvik, John K.; Hægeland, Torbjørn; Raknerud, Arvid

Working Paper

Estimation of Earnings- and Schooling Choice Relations: A Likelihood Approach

Discussion Papers, No. 486

Provided in Cooperation with:

Research Department, Statistics Norway, Oslo

Suggested Citation: Dagsvik, John K.; Hægeland, Torbjørn; Raknerud, Arvid (2006) : Estimation of Earnings- and Schooling Choice Relations: A Likelihood Approach, Discussion Papers, No. 486, Statistics Norway, Research Department, Oslo

This Version is available at:

<https://hdl.handle.net/10419/192468>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Discussion Papers No. 486, November 2006
Statistics Norway, Research Department

*John K. Dagsvik, Torbjørn Hægeland and
Arvid Raknerud*

Estimation of Earnings- and Schooling Choice Relations A Likelihood Approach

Abstract:

In this paper we develop a full information maximum likelihood method for the estimation of a joint model for the choice of length of schooling and the corresponding earnings equation. The model for schooling is assumed to be an ordered probit model, whereas the earnings equation is allowed to be very general with explanatory variables that are flexible transformations of schooling and experience. The coefficients associated with length of schooling and experience are allowed to be random and all the random terms of the model may be correlated. Under normality assumptions, we show that the joint probability distribution for schooling and earnings can be expressed on a closed form that is tractable for empirical analysis.

Keywords: Schooling choice, earnings equation, treatment effects, self-selection, ordered probit, random coefficients, full information maximum likelihood

JEL classification: C31, I20, J30

Acknowledgement: Financial support from The Norwegian Research Council ("KUNI") is gratefully acknowledged.

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no
Torbjørn Hægeland, Statistics Norway, Research Department. E-mail: thd@ssb.no.
Arvid Raknerud, Statistics Norway, Research Department. E-mail: rak@ssb.no,

Discussion Papers

comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
<http://www.ssb.no>
<http://ideas.repec.org/s/ssb/disap.html>

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

The relationship between schooling and earnings is one of the most frequently studied in empirical economics. A large number of these studies build upon versions of the earnings equation proposed by Mincer (1974). A key parameter in the Mincer earnings equation is the coefficient associated with years of schooling. This coefficient is meant to capture the differences in earnings caused by differences in schooling. However, to give a causal interpretation of the parameters of the earnings equation, one must take into account that the independent variable “years of schooling” is endogenous because it is the outcome of a choice variable: level of schooling. The endogeneity problem is related to the fact that the researcher does not observe all factors that affect the schooling choice. For example, if some of these unobservable factors are correlated with unobservables in the earnings equation, OLS will produce biased estimates of the returns to schooling (selection bias).

Traditionally, selection bias is assumed to arise because of correlation between length of schooling and the additive error term in the earnings equation. If such correlation exists and is positive, it implies that people with high earnings capacity (irrespective of level of schooling) would systematically choose a higher schooling level than people with low earnings capacity. Various econometric methods have been developed to deal with this problem, see Griliches (1977) for an overview. In several more recent studies, the coefficient associated with years of schooling is allowed to be individual specific and represented by a random variable in the model. The motivation for this extension is to accommodate heterogeneity in the returns to schooling, that is implied by for example the theoretical model of Willis and Rosen (1979). The random coefficient may also be correlated with the schooling variable and the additive error term in the earnings equation. To deal with this type of endogeneity, two stage control function approaches have often been applied (Heckman, 1979). This approach was introduced in a returns to schooling framework by Garen (1984). There is also a substantial literature on how to interpret instrumental variable estimates in the case of heterogeneity in returns to schooling. See for example Angrist, Imbens and Rubin (1996), Heckman and Vytlačil (1998) and Woodridge (2002). Card (2001) gives an overview of more recent approaches to estimating returns

to schooling in the presence of individual heterogeneity.

The contribution of this paper is to develop a full information maximum likelihood method for the estimation of the parameters of earnings- and schooling decision relations, when we allow for two types of self-selection into schooling: selection by "absolute advantage" (correlation between schooling and the additive error term in the earnings equation) and selection by "comparative advantage" (correlation between schooling and the random coefficients associated with the returns to schooling and experience). In our framework, the decision model for the length of schooling is assumed to be an ordered probit model, whereas the formulation of the earnings equation allows for (generalized) Box-Cox or spline transformations of schooling and experience. A Box-Cox transformation of the dependent variable (earnings) is also allowed for. Under the assumption that the unobserved random components of the model are multnormally distributed, we show that the joint probability distribution for the chosen length of schooling and earnings can be expressed on a closed form that is tractable for empirical analysis.

Compared to the two-stage approach, our maximum likelihood method offers several advantages: First, since estimation is carried out in one stage, we do not have to worry about biased estimates of the standard errors. Such biases may arise because of imputation of parameters estimated in the first first-stage and because, conditional on the individual's choice, the error term is heteroskedastic. Second, since our approach is based on the maximum likelihood method it allows us to deal with non-linear transformations of both earnings, schooling and experience that may contain unknown parameters (such as in Box-Cox transformations). Third, our approach makes it easy to test interesting hypotheses by means of the likelihood ratio test, whereas in the two-stage method exact testing will be cumbersome. Finally, we are able to obtain expressions for the distributions of many random variables of interest, such as the conditional distribution of earnings given the chosen level of schooling. This is useful for assessing various kinds of treatment effects.

A key issue in the recent literature on returns to schooling (and more generally in the program evaluation literature), is the search for valid exclusion restrictions (exogenous variation in the level of schooling) with the purpose of identifying key structural parameters associated with the returns to schooling. We emphasize that also within our framework interpretation of the results depends on the exclusion restrictions, although

such restrictions are not formally needed to obtain identification.

In an application of our method on Norwegian data, it is confirmed that selection effects due to unobservables are important when analyzing the returns to schooling. We find a significant positive correlation between the error term of the schooling choice equation and returns to schooling, and a significant negative correlation between the error term of the schooling choice equation and the additive error term of the earnings equation. Moreover, our study shows, similar to Heckman and Polachek (1974), that for all practical purposes the logarithm of earnings fits the data best (within the class of Box-Cox transformations). Regarding the transformation of the independent variables, we find that piecewise linear functions of “length of schooling” and of “experience” give better fit and also substantially different results than generalized Box-Cox transformations (Box-Cox transformations with arbitrary translations). In particular, we find that the returns to schooling drops markedly after 14 years of schooling. The rest of the paper is organized as follows. In Section 2 we present the modeling framework and derive several results that enable us to carry out empirical inferences. In Section 3 we present the empirical application, while Section 4 concludes the paper.

2 The modelling framework

In this section we specify the modelling framework for estimating the earnings equation and the choice of level of schooling. First consider the relation determining the length of schooling. Let X^* be a latent index that represents the desired level of schooling on a continuous scale. The observed level of education, J , is a categorical variable with M possible categories: $J \in \{1, 2, \dots, M\}$. It is related to X^* through the relation

$$J = j \text{ iff } \mu_{j-1} < X^* < \mu_j, \quad j = 1, \dots, M, \quad (1)$$

where X^* represents the desired level of schooling and $\{\mu_j\}$ are unknown threshold values, except for $\mu_0 = -\infty$ and $\mu_M = \infty$. The variable J represents the choice of level of schooling as constrained by the institutional schooling system, whereas X^* represents the individual’s preferences with regard to the level of schooling on a continuous scale. The threshold values $\{\mu_j\}$ determine the level of schooling in the institutional schooling system

that corresponds to X^* . Furthermore, we assume that

$$X^* = Z_1\gamma_1 + \varepsilon_1, \quad (2)$$

where Z_1 is a row-vector of exogenous variables affecting individual's choice of schooling (typically family background variables describing the situation prior to the choice of schooling), ε_1 is a normally distributed random variable with zero mean and unit variance and γ_1 is a fixed, unknown coefficient vector. Thus, (1)-(2) specifies a standard ordered probit model for the discrete choice variable J .

Consider next the earnings equation. Let $X_1^J(\alpha_1)$ be a transformation of years of schooling and $X_2^J(\alpha_2)$ a transformation of labor market experience. The transformations $X_1^J(\alpha_1)$ and $X_2^J(\alpha_2)$ depend on possibly unknown parameters, α_1 and α_2 , respectively. Each transformation may be a Box-Cox, polynomial, or spline function. The superscript J is used to make explicit that both years of schooling and experience depend on the choice variable, J . For example, for a given age, the maximum level of experience is lower the higher the level of schooling. How $X_1^J(\alpha_1)$ depends on J is determined by the institutional schooling system and can be taken as exogenous: we assume that, conditional on J , the distribution of $X_r^J(\alpha_r)$ is independent of ε_1 as well as the random terms in the earnings relation.

The most general form of our earnings equation is given by

$$(Y^\omega - 1)/\omega = X^J(\alpha)(\beta + \eta) + Z_2\gamma_2 + \varepsilon_2, \quad (3)$$

where ω is an unknown parameter to be estimated, $X^J(\alpha) = (X_1^J(\alpha_1), X_2^J(\alpha_2))$, $\eta = (\eta_1, \eta_2)'$ is a zero mean random coefficient vector, $\beta = (\beta_1, \beta_2)'$ is the corresponding fixed coefficient vector, Z_2 is a vector of exogenous variables which—in addition to variables from Z_1 —also may contain other variables affecting earnings. γ_2 is the corresponding vector of regression coefficients, and ε_2 is a zero mean random term. For any given J , the variation in $X^J(\alpha)$ is assumed to be independent of the error terms $\varepsilon_1, \varepsilon_2$ and η . Note that, with the usual convention that the Box-Cox transformation $(Y^\omega - 1)/\omega = \ln Y$ when $\omega = 0$, the dependent variable in (3) is a continuously differentiable transformation of Y . Also note that, through the random coefficient vector η , our model allows heterogeneity both in the returns to schooling and experience.

Even in the special case where the transformation parameters ω and α are known (or given), one cannot estimate (3) by standard methods due to the fact that $X^j(\alpha)$ depends on ε_1 , which may be correlated with both η and ε_2 . As we show in this paper, it is, however, possible to derive a closed-form expression for the joint probability density of the endogenous variables (Y, J) , given Z_1 and Z_2 .

Let

$$E(\varepsilon_1 \varepsilon_2) = \theta, E(\eta_k \varepsilon_1) = \rho_k. \quad (4)$$

The vector of random terms $(\varepsilon_1, \varepsilon_2, \eta')$ is assumed to be multinormally distributed with zero mean and a general covariance matrix, apart from the conventional identifying restriction that ε_1 has unit variance. In the following, let $\Phi(\cdot)$ denote the standard normal c.d.f. and $\phi(\cdot)$ the corresponding density. We then have the following result.

Theorem 1 *Let Ω be the covariance matrix of (ε_2, η') ,*

$$\psi(X^j(\alpha))^2 = \left[\begin{pmatrix} 1 & X^j(\alpha) \end{pmatrix} \Omega \begin{pmatrix} 1 & X^j(\alpha) \end{pmatrix}' \right] \quad (5)$$

and

$$g(X^j(\alpha))^2 = 1 - \frac{(X^j(\alpha)\rho + \theta)^2}{\psi(X^j(\alpha))^2}.$$

If $f(y, j|Z_1, Z_2)$ denotes the joint density of (Y, J) given (Z_1, Z_2) , then

$$f(y, j|Z_1, Z_2) = \quad (6)$$

$$\begin{aligned} & \frac{y^{\omega-1}}{\psi(X^j(\alpha))} \phi \left(\frac{(y^\omega - 1)/\omega - X^j(\alpha)\beta - Z_2\gamma_2}{\psi(X^j(\alpha))} \right) \times \\ & \left\{ \Phi \left(\frac{\mu_j - Z_1\gamma_1 - b_j((y^\omega - 1)/\omega - X^j(\alpha)\beta - Z_2\gamma_2)}{g(X^j(\alpha))} \right) \right. \\ & \left. - \Phi \left(\frac{\mu_{j-1} - Z_1\gamma_1 - b_j((y^\omega - 1)/\omega - X^j(\alpha)\beta - Z_2\gamma_2)}{g(X^j(\alpha))} \right) \right\}. \end{aligned} \quad (7)$$

The proof of the theorem is given in Appendix A.

The Theorem shows that the joint density of (Y, J) (conditional on Z_1, Z_2) can be expressed by means of the normal c.d.f. and p.d.f.. The first factor in (7) is the marginal distribution of Y when level of schooling is considered as a fixed index (j) – and not as the outcome of a choice variable, J . The second factor expresses the conditional distribution of J given Y .

Next we consider the conditional distribution of η and ε_2 given (J, Z_1, Z_2) . Using (4), we can write

$$\varepsilon_2 = \theta\varepsilon_1 + \tilde{\varepsilon}_2, \eta = \rho\varepsilon_1 + \tilde{\eta}, \quad (8)$$

where $\rho = (\rho_1, \rho_2)'$, and $\tilde{\varepsilon}_2$ and $\tilde{\eta}$ are independent of ε_1 , with

$$\begin{bmatrix} \tilde{\varepsilon}_2 \\ \tilde{\eta} \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$$

for a positive semidefinite 3-dimensional covariance matrix Σ .

Proposition 2 *Let $\kappa(\varepsilon_2|J, Z_1, Z_2)$ and $q_k(\eta_k|J, Z_1, Z_2)$ be the conditional densities of ε_2 and η_k , respectively, given (J, Z_1, Z_2) . If $(\tau_0^2, \tau_1^2, \tau_2^2)$ denotes the diagonal elements of Σ , we have*

$$\kappa(\varepsilon_2|J, Z_1, Z_2) = \frac{1}{b_0} \phi\left(\frac{\varepsilon_2}{b_0}\right) h_{J0}(\varepsilon_2)$$

and

$$q_k(\eta_k|J, Z_1, Z_2) = \frac{1}{b_k} \phi\left(\frac{\eta_k}{b_k}\right) h_{Jk}(\eta_k)$$

where

$$b_0 = \sqrt{\theta^2 + \tau_0^2}, \quad b_k = \sqrt{\rho_k^2 + \tau_k^2}, \quad k = 1, 2,$$

and

$$h_{Jk}(x) = \frac{\Phi\left(\frac{b_k}{\tau_k}(\mu_J - Z_1\gamma_1 - \frac{\rho_k x}{b_k^2})\right) - \Phi\left(\frac{b_k}{\tau_k}(\mu_{J-1} - Z_1\gamma_1 - \frac{\rho_k x}{b_k^2})\right)}{\Phi(\mu_J - Z_1\gamma_1) - \Phi(\mu_{J-1} - Z_1\gamma_1)}, \quad k = 0, 1, 2.$$

The proof of Proposition 2 is given in Appendix A.

The log-likelihood function is given by

$$L = \sum_{i=1}^N \ln f(Y_i, J_i | Z_{1i}, Z_{2i}), \quad (9)$$

where the index i represents *individual* i . The fact that one can express $f(y, j|Z_1, Z_2)$ on closed form has several important advantages. First, it becomes easy to carry out maximum likelihood estimation and to perform statistical tests by means of the likelihood ratio statistic. Second, by utilizing the results in Corollary 3 below, several types of treatment effects, as commonly discussed in the literature, can be estimated. Third, it is easy to extend the model to the case where the random components $(\varepsilon_1, \varepsilon_2, \eta')$ have a mixed multivariate normal distribution, as show in Appendix B. This case is similar to, but much simpler than, Carneiro et al. (2003), whose approach requires that parameters have prior (Bayesian) distributions, and leads to simulation-based inferences.

Corollary 3 *Let $\xi(J) = E(\varepsilon_2|J)$ and $\delta_k(J) = E(\eta_k|J)$. Under the assumptions of Theorem 1,*

$$\delta_k(J) = -\rho_k \lambda(J) \quad (10)$$

and

$$\xi(J) = -\theta \lambda(J), \quad (11)$$

where

$$\lambda(J) = \frac{(\phi(\mu_J - Z_1\gamma_1) - \phi(\mu_{J-1} - Z_1\gamma_1))}{\Phi(\mu_J - Z_1\gamma_1) - \Phi(\mu_{J-1} - Z_1\gamma_1)}. \quad (12)$$

The proof of Corollary 3 is given in Appendix A.

Corollary 3 is useful for calculating the effects of alternative schooling choices. For example, $\delta_k(j)$ yields the mean of η_k for those who have chosen $J = j$. When analyzing the implications of alternative schooling choices, it is of interest to calculate (treatment) effects not only conditional on J , but also conditional on years of schooling, i.e., on $X_1^J = x_1$. Let $\zeta(x_1)$ be the (deterministic function) that assigns the schooling level that corresponds to x_1 years of schooling, i.e., $X_1^J = x_1$ implies $J = \zeta(x_1)$. Thus, the following holds

$$E(\eta_k|X_1^J = x_1) = \delta_k(\zeta(x_1)) \quad (13)$$

and

$$E(\varepsilon_2|X_1^J = x_1) = \xi(\zeta(x_1)). \quad (14)$$

Note that we can express (3) as

$$(Y^\omega - 1)/\omega = X^J(\alpha)(\beta + \delta(J)) + Z_2\gamma_2 + \xi(J) + \varepsilon^*,$$

where

$$\varepsilon^* = \varepsilon_2 - \xi(J) + X^J(\alpha)(\eta - \delta(J)),$$

$\delta(J) = (\delta_1(J), \delta_2(J))$, and the error term, ε^* , has the property that $E(\varepsilon^*|J) = 0$. Thus, it is possible to estimate β and γ_2 by (possibly non-linear) regression, because the mean of the error term given the self-selected sample is zero and estimates of $\delta(J)$ and $\xi(J)$ are available from (10)-(12). Specifically, a conventional ordered probit analysis based on (1)-(2) yields estimates of μ_J and γ_1 .

The special case with $\omega = 0$ and with no transformation of years of schooling, and experience expressed as a quadratic, corresponds to the standard Mincer equation that has been used in numerous empirical studies on "returns to schooling". Most of the papers explicitly addressing selection bias in returns to schooling either use instrumental variables or a version of the two-stage method outlined above; see, for example, Card (2000) and Vella and Verbeek (1999).

3 An empirical application

3.1 Data

The data for this application are taken from the Norwegian system of register data, where individual information about essentially all Norwegian residents is gathered from a number of governmental administrative registers. In addition to basic demographic information, the system contains information about education, income and employment. In this study, we use a 10 percent sample of all native-born males who lived in Norway in 1970, born between 1952 and 1970, and still living in Norway in 1997. The data contain information on years of schooling and type of education for each individual. The earnings equation sample is restricted to full-time wage-earners, defined as individuals working 30 hours or more per week, leaving us with 29,533 observations. Labor market experience is represented in the usual way, i.e., age minus years of schooling minus seven years. The earnings measure used is total annual taxable labor income. Because the earnings measure reflects annual earnings, observations where employment relationships started or terminated within the actual year were excluded. Holders of multiple jobs and individuals

who have received labor market compensation or have participated in active labor market programs have been excluded. Family background information is taken from the National Census of the Population and Housing in 1970. A full list of variables with key summary statistics is given in Table A1.

3.2 Empirical specification and estimation results

In our application the level of schooling is divided into eight groups, i.e., $J \in \{1, 2, \dots, 8\}$. Level 1 corresponds to seven to nine years of schooling, levels 2 to 7 correspond to 10-15 years, respectively, whereas level 8 corresponds to 16-18 years of schooling. Let X_1^J denote years of schooling *exceeding* seven years and X_2^J potential experience, defined as age minus years of schooling minus seven years. We consider three types of transformation functions of X_k^J ($k = 1, 2$), namely

$$\begin{aligned} \text{Linear:} & X_k^J(\alpha_k) = X_k^J \\ \text{Quadratic:} & X_k^J(\alpha_k) = [(X_k^J + \alpha_{k,1})^2 - 1] / 2 \\ \text{Generalized Box-Cox:} & X_k^J(\alpha_k) = [(X_k^J + \alpha_{k,1})^{\alpha_{k,2}} - 1] / \alpha_{k,2} \\ \text{Splines:} & X_k^J(\alpha_k) = \sum_j \alpha_{k,[j/2]}, \alpha_{k,0} = 1, \end{aligned}$$

where $[x]$ denotes the integer value of x . The spline transformation of X_k^J has knots at year one and two and every second year thereafter (4,6,8,...). Thus, because the maximum values of X_1^J and X_2^J in our sample is 11 (i.e., 18 years of schooling) and 29 (years of experience), respectively, we are able to identify five α_1 -parameters ($[11/2] = 5$) and 14 α_2 -parameters ($[29/2] = 14$). Note that the linear and quadratic transformations are special cases of the (generalized) Box-Cox transformation, obtained by setting $\alpha_{k,2} = 1$ and $\alpha_{k,2} = 2$, respectively.

The vector of explanatory variables in the income equation, Z_2 , includes indicators about sector of occupation (public, private services, manufacturing), field of education (general, technical, humanistic, teaching, administrative, etc.), and indicators for each of 19 counties where the individual works. The vector of explanatory variables of the ordered probit model for schooling choice, Z_1 , contains variables regarding the family background. These include dummy variables for birth cohort, indicators of whether the individual as a child lived with both parents or alone with either mother or father, the labor market status of the parents, indicators of household income (quintile and both the father's and mother's education level), and whether the person had a mother and/or father who was

born abroad. In addition, the schooling choice equation contains indicator variables for the county where the individual grew up, for example, where the individual lived in 1970. The main exclusion restriction in this application, which in addition to functional form assumptions identifies the parameters of the model, is that given all the other covariates in the model, the region where you grew up may affect your choice of schooling, but not your earnings. It is well documented that educational choices vary considerably across regions in Norway. This is true also when conditioning on, for example, family background variables. The instrument is in the spirit of Card (1995) who used college proximity as an instrument, but may be interpreted in a more general sense as variations in the opportunity cost of education.

The results for some key combinations of transformations of earnings, schooling and experience are displayed in Table 1. A full set of results is reported in the Appendix C. When interpreting the results in the table, one should bear in mind that the parameter estimates of β_1 and β_2 are not comparable across different models, as they are coefficients of different transformations of schooling and experience. Moreover, whereas the models reported in the first three columns of Table 1 have log income, $\ln y$, as the dependent variable, the last column reports results from a specification with a general Box-Cox transformation of income.

From Table 1, we first note that the linear-quadratic specification with regard to schooling and experience, i.e., the traditional Mincer model, gives a substantially lower log-likelihood than the Box-Cox model (Model 2) and – especially – the spline models (models 3-4). On the other hand, when $\omega = 0$, the spline transformations of X_1^J and X_2^J give considerably higher likelihood than the Box-Cox transformations – but at the cost of 15 more parameters. Although the model with spline transformations of X_1^J and X_2^J is clearly the most flexible one with respect to parameterization, it is not a special case of either the Box-Cox or the linear-quadratic specification. On the other hand, the linear-quadratic specification is a special case of Box-Cox, with three parameters fewer. Because $\hat{\alpha}_{2,1} = 2.49$ and $\hat{\alpha}_{2,2} \approx 0$ (the “hat” denotes a maximum likelihood estimate), we see that the estimated Box-Cox transformation of experience amounts to $\ln(X_2^J + 2.49)$.

With regard to the transformation of income, the general Box-Cox transformation leads to an estimate of ω equal to -.17, with a standard error of only .003. The results

Table 1: **Parameter estimates of earnings equation for different model specifications**

	Model specification			
	Model 1	Model 2	Model 3	Model 4
Earnings:	$\ln y$	$\ln y$	$\ln y$	$(Y^\omega - 1)/\omega$
Schooling:	Linear	Box-Cox	Splines	Splines
Experience:	Quadratic	Box-Cox	Splines	Splines
ω	0 (-)	0 (-)	0 (-)	-.17 (.003)
β_1	.08 (.003)	.66 (.09)	.06 (.005)	.01 (.001)
β_2	-.0005 (.0001)	.21 (.01)	.01 (.001)	.05 (.005)
$SD(\varepsilon_2)$.26 (.01)	.51 (.10)	.24 (.03)	.05 (.001)
$SD(\eta_1)$.01 (.004)	.08 (.02)	.01 (.001)	.001 (.0002)
$SD(\eta_2)$.0003 (.00004)	.13 (.02)	.01 (.001)	.03 (.004)
$\text{Corr}(\varepsilon_2, \varepsilon_1)$	-.12 (.05)	-.34 (.06)	-.25 (.05)	-.25 (.04)
$\text{Corr}(\eta_1, \varepsilon_1)$.03 (.24)	.39 (.06)	.35 (.07)	.47 (.16)
$\text{Corr}(\eta_1, \varepsilon_2)$.92 (.13)	-.82 (.06)	-.13 (.23)	.30 (.16)
$\text{Corr}(\eta_2, \varepsilon_1)$	-.04 (.07)	-.01 (.03)	-.03 (.03)	.06 (.05)
$\text{Corr}(\eta_2, \varepsilon_2)$	-.62 (.07)	-.93 (.02)	-.48 (.11)	-.15 (.08)
$\text{Corr}(\eta_2, \eta_1)$	-.87 (.17)	.77 (.06)	.64 (.11)	-.77 (.15)
$\alpha_{1,1}$		11.26 (1.39)		
$\alpha_{1,2}$	1 (-)	.28 (.03)		
$\alpha_{2,1}$	-29.5 (2.48)	2.49 (.62)		
$\alpha_{2,2}$	2 (-)	.002 (.002)		
log-likelihood	-27274	-27251	-27179	-27058
Sample size	29,332	29,332	29,332	29,332

suggest that ω is significantly different from zero. However, from the point of view of economic significance $\omega = -.17$ is so close to zero that the Box-Cox and logarithmic transformation are equivalent for practical purposes. We will illustrate this point below.

The estimated correlations between the stochastic terms have interesting economic interpretations and give information on the nature of self-selection. However, the pairwise correlations reported in Table 1 show that many of these are not robust across different model specifications. For example, we find strong evidence of negative correlation between η_2 and ε_2 when $\omega = 0$, but not at the maximum likelihood estimate $\omega = -.17$. However, with regard to the correlations that have the clearest economic interpretation we get quite striking results. First of all, it is evident that self-selection does matter. Concentrating on the results from the Box-Cox and spline transformations of schooling and experience, which overall give the best fit to the data and the most plausible results,

there are significant negative correlations between ε_1 and ε_2 , i.e., the residual terms of the earnings and schooling equations. We also find strong positive correlations between η_1 and ε_1 . Using spline transformations of X_1^J and X_2^J , we obtain correlation coefficients of the same magnitude as for the Box-Cox transformations, regardless of whether $\omega = 0$ or $\omega = -.17$. The robust findings that $\text{Corr}(\eta_1, \varepsilon_1) > 0$ and $\text{Corr}(\varepsilon_1, \varepsilon_2) < 0$ imply, respectively, that individuals who undertake more education than what is predicted from the schooling equation, have high returns to schooling, and that the part of their earnings potential that is unrelated to schooling and experience is lower. In particular, the economic interpretation of $\text{Corr}(\eta_1, \varepsilon_1) > 0$ is that individuals with a high learning potential in school also have a high learning potential on the job. It should be kept in mind, however, that the correlations reported in Table 1 depend on the respective specifications and cannot be interpreted independently of the chosen transformations of length of schooling and experience.

There is considerable heterogeneity in the returns to schooling and experience, as seen from the estimated standard deviations $\text{SD}(\eta_1)$ and $\text{SD}(\eta_2)$ of η_1 and η_2 , respectively, which are of the same magnitude as the estimated fixed coefficients, $\hat{\beta}_1$ and $\hat{\beta}_2$. To evaluate the importance of individual heterogeneity in the returns to experience and schooling, it is natural to look at the variation coefficients $\text{SD}(\eta_1)/\hat{\beta}_1$ and $\text{SD}(\eta_2)/\hat{\beta}_2$. These ratios lie between 1/10 and 1 in all the model specifications and are smaller for schooling than for experience. Thus, it seems that relative to the fixed coefficient, β_k , the unobserved heterogeneity in returns to experience is larger than in returns to schooling. As a further check of the importance of heterogeneity in the coefficients of schooling and experience, a model with only a fixed coefficient vector (i.e., no η -vector) was estimated. This restriction reduced the number of parameters by nine. However, it was firmly rejected by a likelihood ratio tests.

In analyses of returns to schooling and experience, the marginal returns to schooling and the earnings-experience profiles are of key interest. In models allowing for heterogeneity in returns, there are several possible “marginal returns” or “treatment effects” that may be calculated, based on the estimation results. Which effects that are most relevant, depend on the purpose of the analysis. In models with no heterogeneity in the returns, all treatment effects coincide.

The differences in results across the four model specifications are illustrated in Figures 1–4, along with the results from a linear-quadratic specification without selection effects (equivalent to OLS estimation of a standard Mincer equation). Figure 1 shows expected log earnings as a function of years of schooling when all the other variables of the earnings equation are set equal to their sample mean. In particular, years of experience is fixed at 15 years. The intercepts of the different graphs in the figure are determined by the (identifying) condition that when all the variables are at their sample means, expected log earnings should be equal in all the four model specifications. We see that the two versions of the model with spline transformations of schooling depicted in Figure 1, i.e., with $\omega = 0$ and $\omega = -.17$ as the dependent variables, are almost identical, except for small discrepancies at low values of years of schooling.

Figure 2 shows the expected marginal returns to schooling corresponding to the three specifications in Figure 1 with $\ln Y$ as the dependent variable. The natural interpretation of the estimates from the models with selection effects is as the “average treatment effect” of schooling. This means that the graphs show the marginal effect on earnings of the *last* year of schooling, given that a randomly selected individual is given that number of years of schooling (years of schooling is shown on the horizontal axis). In contrast, the interpretation of the OLS estimate shows the (conditional) earnings differentials between individuals with different levels of schooling. In the absence of selection effects, OLS and the linear specification will coincide.

Comparing OLS with full information maximum likelihood estimation of the linear specification, we see from Figures 1 and 2 that allowing for selection effects does matter for the estimated returns to schooling. From Figure 2 we find a marginal returns to schooling of around one percentage point higher when we allow for selection effects. When comparing the linear specification with the more flexible specifications, we see that there are considerable differences in the estimated marginal returns across levels of schooling. In particular, there are high returns to completing upper secondary school (12 years) and to take one or two years of higher education, while the marginal return of the last year of schooling, if the current level of schooling is 15 years or more, is considerably smaller. Hence, allowing for a more flexible specification not only improves the fit of the model as measured by log-likelihood, but also affects the size of the key measures of interest.

Figure 3 shows expected log earnings as a function of years of experience, with the other variables of the earnings equation fixed at their sample means, for example, years of schooling equals 12 years. In contrast to returns to schooling, allowing for selection effects only has minor implications for estimated returns to experience. We see from Figure 3 that the Box-Cox specification gives higher marginal returns for years of experience up to four to five years compared to the other specifications.

Concentrating on our preferred specification, with spline transformations of both schooling and experience and with log earnings, $\ln Y$, as the dependent variable in the earnings equation, Figure 4 depicts three different kinds of marginal returns to schooling. The first is the average treatment effect (ATE), $\beta_1 \Delta x_1(\alpha_1)$, that was also depicted in Figure 2, where $\Delta x_1(\alpha_1)$ is the change in the spline transformation of schooling, $x_1(\alpha_1)$, when years of schooling increases from $x_1 - 1$ to x_1 . The second is the “effect of the treatment on the treated” (TT), $(\beta_1 + \delta_1(\zeta(x_1 - 1))) \Delta x_1(\alpha)$, cf. (13), which has the interpretation of the marginal return by increasing years of schooling from $x_1 - 1$ to x_1 for those who did in fact undertake the investment. The final is the observed differentials between levels of schooling (OD): $(\beta_1 + \delta_1(\zeta(x_1 - 1))) \Delta x_1(\alpha) + \Delta \xi(\zeta(x_1))$, where $\Delta \xi(\zeta(x_1)) \equiv \xi(\zeta(x_1)) - \xi(\zeta(x_1 - 1))$, cf. (14). This is the sum of (i) the average treatment effect, (ii) the average of the idiosyncratic marginal returns to schooling for the individuals with this level of schooling and (iii) the average idiosyncratic earnings level effect for the same individuals. We see that TT in general is higher than ATE. This reflects the positive correlation between ε_2 and η_1 that was reported in Table 1: Individuals with higher idiosyncratic return to a level of schooling also invest more in schooling. Hence the marginal returns at a specific level are higher for those who actually have this level of schooling than for the average individual. In other words, there is selection by comparative advantage. On the other hand, we also estimated a negative correlation between ε_1 and ε_2 ; conditional on idiosyncratic returns to schooling, those with higher earnings potential regardless of schooling tend to choose a lower level of schooling. This is clearly seen from the earnings-schooling profiles in Figure 1. The self-selection related to ε_2 gives a flatter profile, i.e., individuals with high ε_2 tend to have low levels of schooling and vice versa.

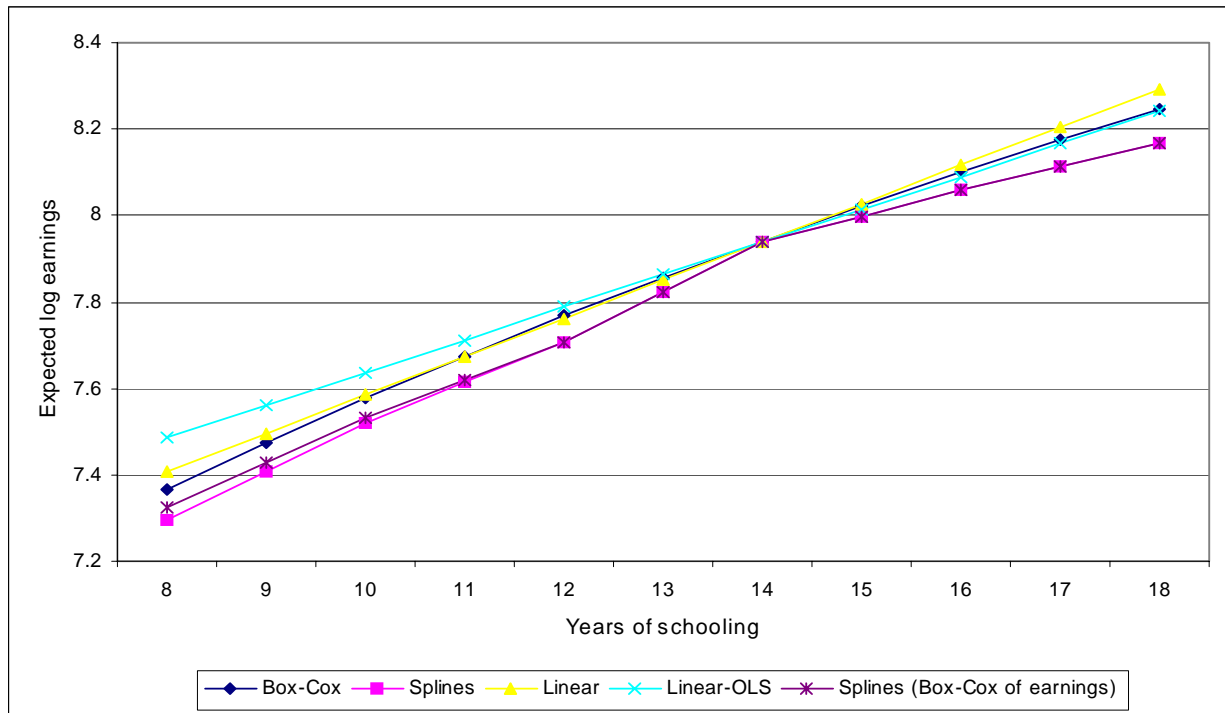


Figure 1: Expected log earnings as a function of years of schooling

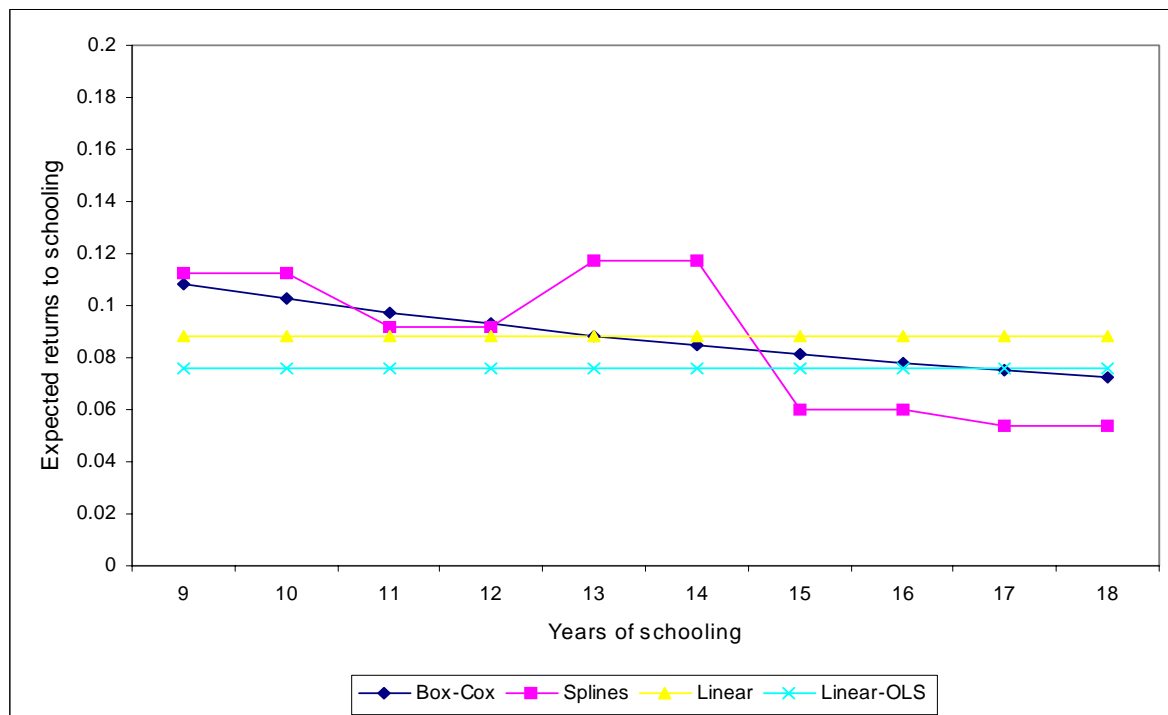


Figure 2: Expected marginal returns to schooling

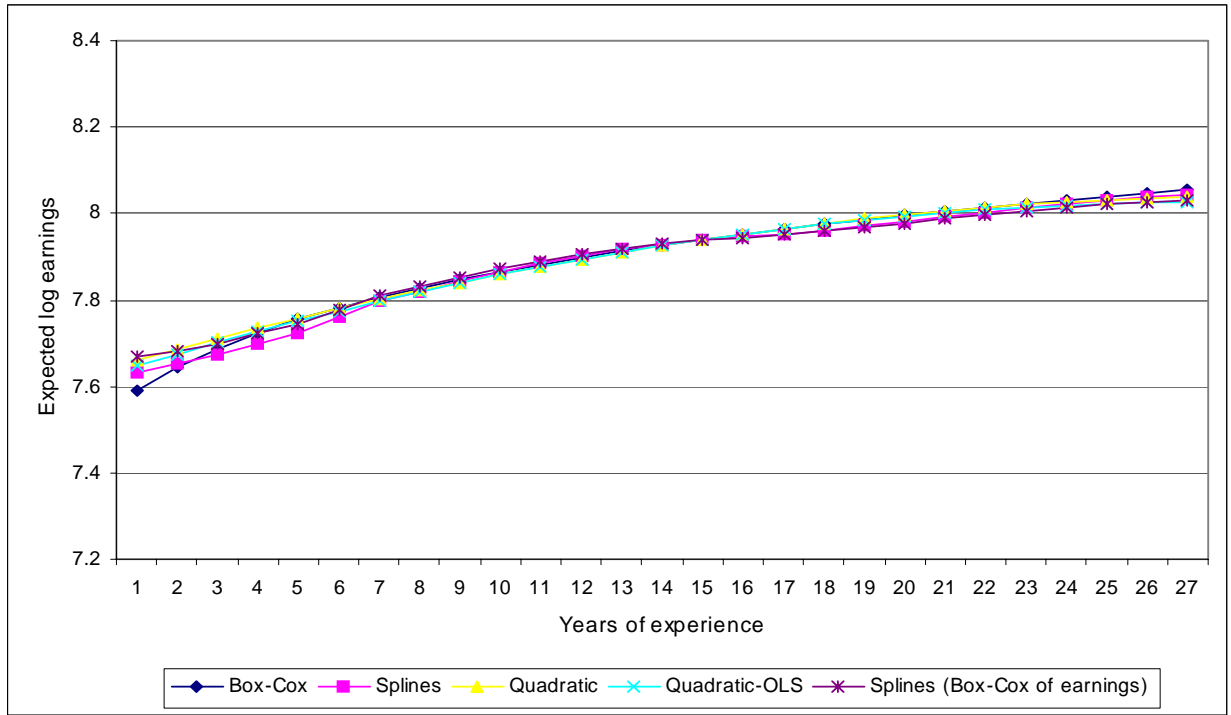


Figure 3: Expected log earnings as a function of experience

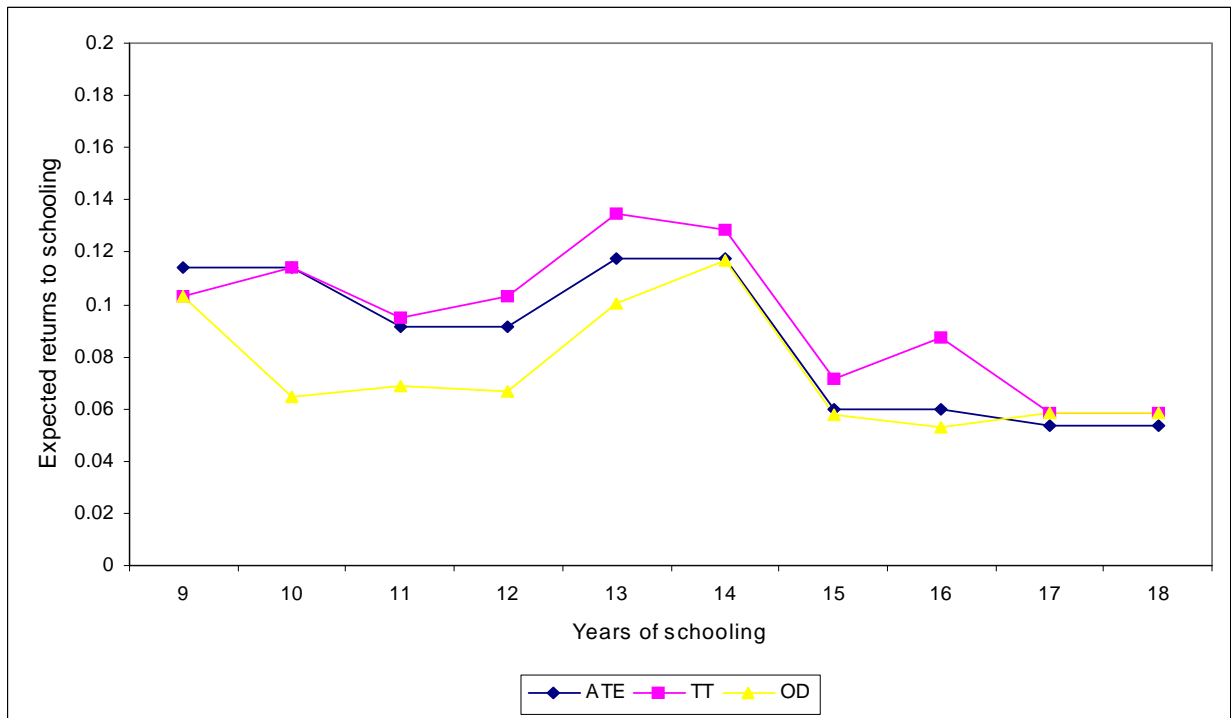


Figure 4: Estimates of treatment effects

To evaluate the fit of our preferred specification, Figure 5 plots (i) the discrete prob-

ability density functions over a grid of 100 intervals, with equal length, for the estimated spline model with log earnings, $\ln Y$, as the dependent variable, and (ii) histograms of the log earnings data. This is done conditional on the chosen level of schooling, i.e., for eight different levels. Note that the estimated theoretical models are not normal distributions. They are derived from (7), by integrating out (Z_1, Z_2) using the empirical distribution function of these covariates. In fact, the estimated conditional probability density functions are slightly skewed to the right, although this is barely visible in the figure. A QQ-plot of the marginal distribution of log-earnings is presented in Figure 6. The overall impression from these figures is that the estimated model fits the data well.

4 Conclusion

In this paper we have discussed maximum likelihood estimation of a joint model for earnings and choice of level of schooling. The earnings relation is allowed to be very general with random coefficients and possibly particular families of nonlinear transformations of the independent and dependent variables. The choice model for length of schooling is an ordered Probit model. The random coefficients and the additive error terms in the earnings relation and the choice model are assumed distributed according to the multinormal distribution. This means that all the random terms in the model are allowed to be correlated. Under these assumptions we have demonstrated that the joint distribution of the choice of level of schooling and earnings can be expressed on closed form. We have also outlined how this model can be extended to the case where the joint distribution of the random terms is a discrete (multinomial) mixture of the multinormal distribution.

We have applied this framework and methodology to analyze the structure of the earnings relation on microdata for Norway. The estimation results show that if we constrain the transformation of the dependent variable to be of the Box-Cox type, the logarithm of earnings seems to be the best one in terms of fit. Within the class of Box-Cox transformations, or alternatively spline transformations of the independent variables “years of schooling” and “potential experience”, the latter family turns out to give the best fit.

We believe that the econometric framework developed in this paper offers several advantages to the researcher compared to the two-stage control function approach. Because it is a maximum likelihood approach it allows for nonlinear transformations of the depen-

dent variable that contain unknown parameters. Second, biases due to heteroscedasticity and imputed estimates from the first stage that typically plague the control function approach no longer exist. Third, the maximum likelihood approach facilitates testing of alternative specifications within our framework.

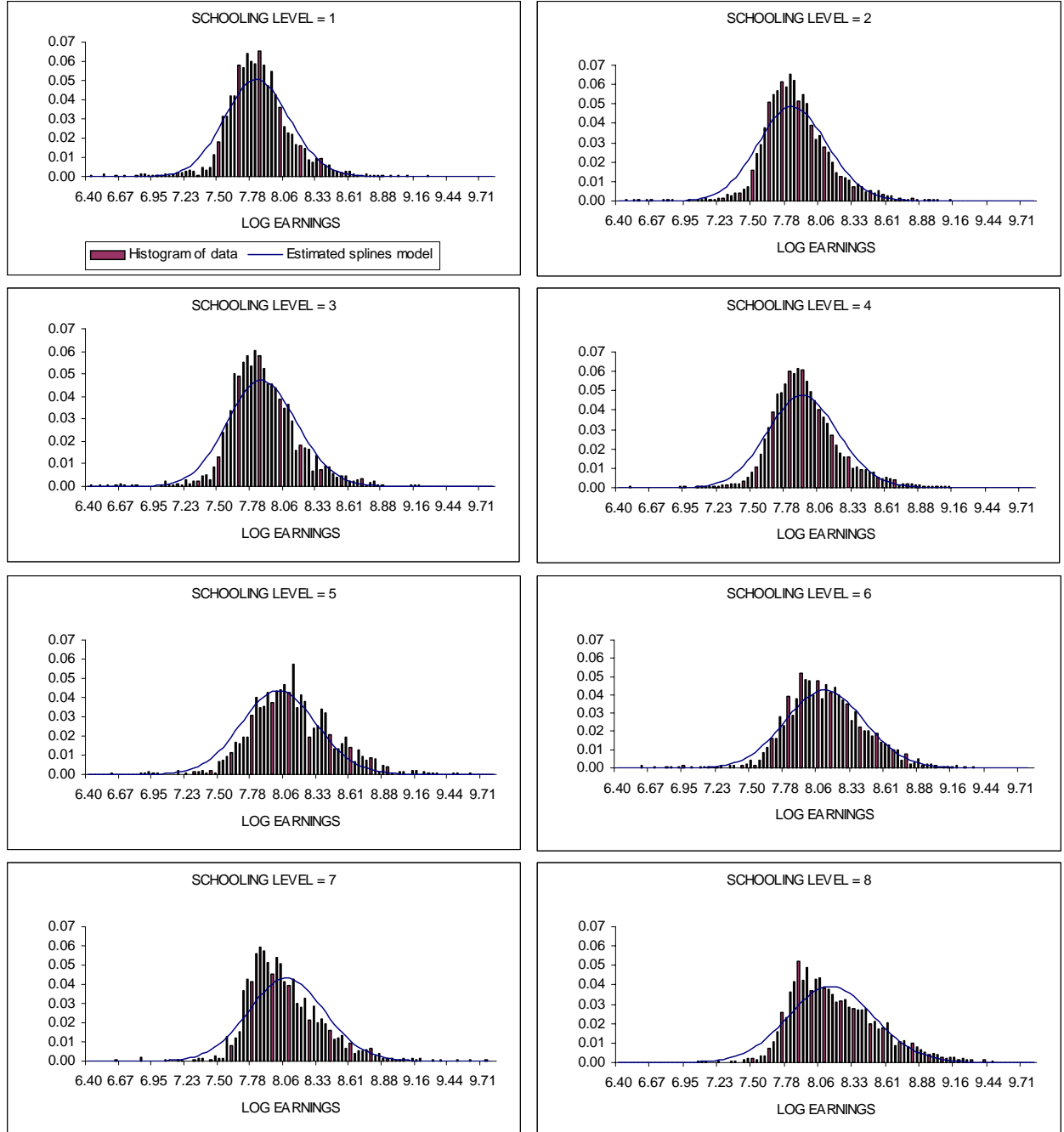


Figure 5: Comparing estimated probability distribution functions and histograms of log earning data: The conditional distributions of log earnings given the level of schooling

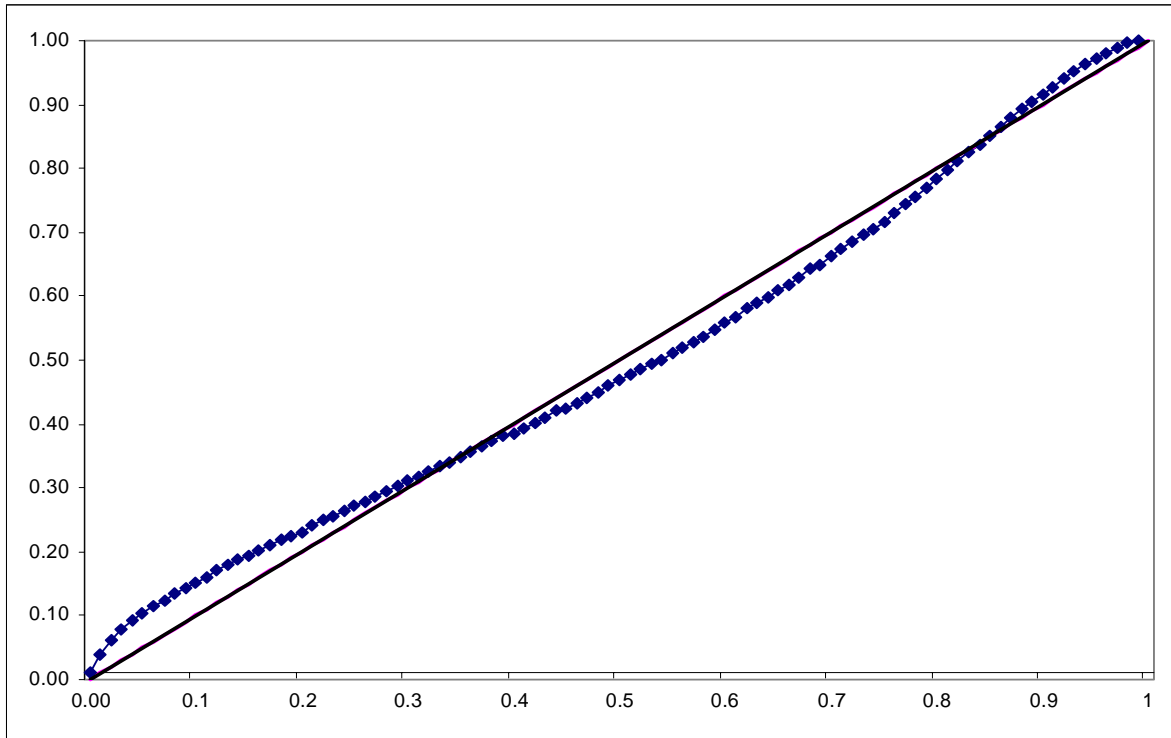


Figure 6: Comparing estimated and empirical distribution functions: QQ-plot for the marginal distribution of log earnings

References

- [1] Angrist, J.D., G.W Imbens and D.B. Rubin (1996): Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*. 91, 444–472.
- [2] Card, D. (1995): Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in L.N. Christofides, E.K. Grant and R. Swidinsky (eds.): *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.
- [3] Card, D. (2001): Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69, 1127–1160
- [4] Carneiro, P., Hansen, K. and Heckman, J.J. (2003): Estimating Distributions of Treatment Effects with an Application to the Return to Schooling and Measurement

- of the Effects of Uncertainty on College Choice. *International Economic Review*, 44, 361–422.
- [5] Garen, J. (1984): The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica*, 52, 1199–1218.
 - [6] Griliches, Z. (1977): Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45, 1–22.
 - [7] Gronau, R. (1974): Wage Comparisons – A Selectivity Bias. *Journal of Political Economy*, 82, 1119–1143.
 - [8] Heckman, J.J. (1974): Shadow Prices, Market Wage, and Labor Supply. *Econometrica*, 42, 679–694.
 - [9] Heckman, J.J. (1979): Sample Selection Bias as a Specification Error. *Econometrica*, 47, 153–162.
 - [10] Heckman, J.J. and Polachek, S. (1974): Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship. *Journal of the American Statistical Association*, 69, 350–354.
 - [11] Heckman, J.J. and Vytlacil, E. (1998): Instrumental Variable Methods for the Correlated Random Coefficient model. *Journal of Human Resources*, 33, 974–987.
 - [12] Mincer, J. (1974): *Schooling, Experience and Earnings*. New York: Columbia University Press.
 - [13] Vella, F. and M. Verbeek (1999): Estimating and Interpreting Models with Endogenous Treatment Effects. *Journal of Business and Economic Statistics*, 17, 473–478.
 - [14] Willis, R.J. and S. Rosen (1979): Education and Self-Selection. *Journal of Political Economy*, 87, S7–S36.
 - [15] Wooldridge, J.M. (1997): On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model. *Economic Letters*, 56, 129–133.

- [16] Wooldridge, J.M. (2002): Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model. *Economic Letters*, 79, 185–191.

Appendix A: Proof of Theorem 1, Proposition 2 and Corollary 3

PROOF OF THEOREM 1.

Define

$$\varepsilon_{j3} = X^j(\alpha)\eta + \varepsilon_2. \quad (15)$$

Then we can write

$$\varepsilon_1 = b_j\varepsilon_{j3} + \tilde{\varepsilon}_{j3}, \quad (16)$$

for a suitable fixed coefficient b_j , where $\tilde{\varepsilon}_{j3}$ is normally distributed with zero mean and independent of ε_{j3} . From (15) and (5) it follows that

$$Var(\varepsilon_{j3}) = \psi(X^j(\alpha))^2. \quad (17)$$

By multiplying (16) by ε_{j3} and taking expectation on both sides we obtain that b_j is determined by

$$\begin{aligned} E(\varepsilon_1\varepsilon_{j3}) &= E(b_j\varepsilon_{j3}^2) + E(\tilde{\varepsilon}_{j3}\varepsilon_{j3}) \\ &\Downarrow \\ (X^j(\alpha)\rho + \theta) &= b_j\psi(X^j(\alpha))^2, \end{aligned} \quad (18)$$

where we have used (4), (15) and (17). Moreover, (16) and (17) imply that

$$1 = b_j^2\psi(X^j(\alpha))^2 + Var(\tilde{\varepsilon}_{j3}) \quad (19)$$

When b_j , determined by (18), is inserted into (19), we obtain

$$Var(\tilde{\varepsilon}_{j3}) \equiv g(X^j(\alpha))^2 = 1 - \frac{(X^j(\alpha)\rho + \theta)^2}{\psi(X^j(\alpha))^2}. \quad (20)$$

Now consider the choice of level of schooling. From (1)-(2) and (16),

$$J = j \Leftrightarrow \mu_{j-1} - Z_1\gamma_1 - b_j\varepsilon_{j3} < \tilde{\varepsilon}_{j3} \leq \mu_j - Z_1\gamma_1 - b_j\varepsilon_{j3}. \quad (21)$$

From (3),

$$\varepsilon_{j3} = Y^{[\omega]} - X^j(\alpha)\beta - Z_2\gamma_2. \quad (22)$$

where $Y^{[\omega]} = (Y^\omega - 1)/\omega$. Hence,

$$\begin{aligned} & P(Y^{[\omega]} \in (z, z + dz), J = j | Z_1, Z_2) = \\ & P(Y^{[\omega]} \in (z, z + dz), \mu_{j-1} - Z_1\gamma_1 - b_j\varepsilon_{j3} < \tilde{\varepsilon}_{j3} \leq \mu_j - Z_1\gamma_1 - b_j\varepsilon_{j3}) = \\ & P(\varepsilon_{j3} \in (z - X^j(\alpha)\beta + Z_2\gamma_2, z - X^j(\alpha)\beta + Z_2\gamma_2 + dz), \mu_{j-1} - Z_1\gamma_1 \\ & - b_j[z - X^j(\alpha)\beta + Z_2\gamma_2] < \tilde{\varepsilon}_{j3} \leq \mu_j - Z_1\gamma_1 - b_j[z - X^j(\alpha)\beta + Z_2\gamma_2]) . \end{aligned} \quad (23)$$

Because $\tilde{\varepsilon}_{j3}$ and ε_{j3} are independent, using (17), (20), (7), the last expression above is equal to

$$\begin{aligned} & P(\varepsilon_{j3} \in (z - X^j(\alpha)\beta + Z_2\gamma_2, z - X^j(\alpha)\beta + Z_2\gamma_2 + dz) \times \\ & P(\mu_{j-1} - Z_1\gamma_1 - b_j[z - X^j(\alpha)\beta + Z_2\gamma_2] < \tilde{\varepsilon}_{j3} \leq \mu_j - Z_1\gamma_1 \\ & - b_j[z - X^j(\alpha)\beta + Z_2\gamma_2]) = \\ & \frac{dz}{\psi(X^j(\alpha))} \phi\left(\frac{z - X^j(\alpha)\beta - Z_2\gamma_2}{\psi(X^j(\alpha))}\right) \\ & \left\{ \Phi\left(\frac{\mu_j - Z_1\gamma_1 - b_j(z - X^j(\alpha)\beta - Z_2\gamma_2)}{g(X^j(\alpha))}\right) \right. \\ & \left. - \Phi\left(\frac{\mu_{j-1} - Z_1\gamma_1 - b_j(z - X^j(\alpha)\beta - Z_2\gamma_2)}{g(X^j(\alpha))}\right) \right\} . \end{aligned}$$

Now, because $z = (y^\omega - 1)/\omega$ and $dz = y^{\omega-1}dy$, the density in terms of untransformed earnings, y , becomes equal to (7). This completes the proof. ■

PROOF OF PROPOSITION 2.

From (8) we have, for $k = 1, 2$,

$$q_k(\eta_k | j, Z_1, Z_2) = \frac{\int_{\mu_{j-1} - Z_1\gamma_1}^{\mu_j - Z_1\gamma_1} \frac{1}{\tau_k} \phi\left(\frac{\eta_k - \rho_k \varepsilon_1}{\tau_k}\right) \phi(\varepsilon_1) d\varepsilon_1}{P(J = j)}$$

Because

$$\phi\left(\frac{\eta_k - \rho_k \varepsilon_1}{\tau_k}\right) \phi(\varepsilon_1) = \phi\left(\frac{\eta_k}{b_k}\right) \phi\left(\frac{b_k}{\tau_k}(\varepsilon_1 - \frac{\rho_k \eta_k}{b_k^2})\right),$$

with

$$b_k = \sqrt{\rho_k^2 + \tau_k^2},$$

and

$$\begin{aligned} & \int_{\mu_{j-1} - Z_1 \gamma_1}^{\mu_j - Z_1 \gamma_1} \phi\left(\frac{b_k}{\tau_k}(\varepsilon_1 - \frac{\rho_k \eta_k}{b_k^2})\right) d\varepsilon_1 \\ &= \frac{\tau_k}{b_k} \left(\Phi\left(\frac{b_k}{\tau_k}(\mu_j - Z_1 \gamma_1 - \frac{\rho_k \eta_k}{b_k^2})\right) - \Phi\left(\frac{b_k}{\tau_k}(\mu_{j-1} - Z_1 \gamma_1 - \frac{\rho_k \eta_k}{b_k^2})\right) \right), \end{aligned}$$

we obtain

$$q_k(\eta_k | j, Z_1, Z_2) = \frac{1}{b_k} \phi\left(\frac{\eta_k}{b_k}\right) h_{jk}(\eta_k).$$

Notice that $\frac{1}{b_k} \phi\left(\frac{\eta_k}{b_k}\right)$ is the unconditional distribution of η_k . When $\rho_k = 0$, the correction factor $h_{jk}(x) = 1$. The result for $\kappa(\varepsilon_2 | j, Z_1, Z_2)$ follows similarly. ■

PROOF OF COROLLARY 3.

From (8), (12) and the independence of ε_1 and $\tilde{\eta}$ it follows that

$$\begin{aligned} \delta(j) &= E(\eta | J = j) = E(\rho \varepsilon_1 + \tilde{\eta} | J = j) = E(\rho \varepsilon_1 | J = j) \\ &= \rho E(\varepsilon_1 | \mu_{j-1} - Z_1 \gamma_1 < \varepsilon_1 \leq \mu_j - Z_1 \gamma_1) = \rho \frac{\int_{\mu_{j-1} - Z_1 \gamma_1}^{\mu_j - Z_1 \gamma_1} u \phi(u) du}{P(J = j)} \\ &= -\rho \frac{(\phi(\mu_j - Z_1 \gamma_1) - \phi(\mu_{j-1} - Z_1 \gamma_1))}{\Phi(\mu_j - Z_1 \gamma_1) - \Phi(\mu_{j-1} - Z_1 \gamma_1)} = -\rho \lambda_j. \end{aligned} \tag{24}$$

Similarly, we obtain

$$\begin{aligned} \xi(j) &= E(\varepsilon_2 | J = j) = E(\theta \varepsilon_1 + \tilde{\varepsilon}_2 | J = j) \\ &= \theta E(\varepsilon_1 | J = j) = -\theta \lambda_j. \end{aligned} \tag{25}$$

Appendix B: Extension to mixture distributions of the random components

Similarly to Carneiro et al. (2003) we shall now consider the case where the joint distribution of the random components is a discrete mixture of multinormal distributions. To this end we extend the previous notation of the random error terms to $(\varepsilon_1(R), \varepsilon_2(R), \eta(R))$,

where R denotes a random index that is multinomially distributed with $P(R = r) = p_r$, $r = 1, 2, \dots, Q$. The vectors $(\varepsilon_1(r), \varepsilon_2(r), \eta(r))$, $r = 1, 2, \dots, Q$, are independent and multinormally distributed with zero mean and covariance matrix $\Omega(r)$. With this extension, our model becomes a (non-Bayesian) version of the model estimated in Carneiro et al. (2003) (see Section 7 and Appendix B in their paper). Their model is more general in the sense that they have several measurement and outcome equations in addition to the schooling choice equation, whereas we only have one outcome equation, namely the earnings equation. Now let $f(y, j|R, Z_1, Z_2)$ denote the conditional density of earnings, Y , and chosen schooling level, J , given (R, Z_1, Z_2) . This density is expressed in (7), apart from the modification that the parameters of the covariance matrix $\Omega(r)$, are now indexed by r , i.e., $\theta(r), \rho(r), \Omega(r)$. Consequently, the joint density of (Y, J) can be expressed as

$$f(y, j|Z_1, Z_2) = \sum_{r=1}^Q p_r f(y, j|r, Z_1, Z_2)$$

Thus, the likelihood function can be expressed on closed form also in the case when the joint distribution of the random components is a discrete mixture of multinormal distributions.

Appendix C: Descriptive statistics

Table 2: **Descriptive statistics for schooling, experience and earnings**

Variable	Mean	St.dev.	Min	Max
Years of schooling	12.2	2.3	7	18.0
Years of experience	15.2	5.9	0	29.0
Log of earnings	7.7	0.3	6.4	9.8

Table 3: Descriptive statistics for Z_1 with corresponding parameter estimates

Z1-variables	Mean	St.dev	Min	Max	Parameter estimate	S.E.
Lone mother	0.05	0.21	0	1	0.27	0.07
Lone father	0.01	0.08	0	1	-0.05	0.08
No parents	0.01	0.09	0	1	-0.13	0.07
Mother working	0.33	0.47	0	1	0.02	0.01
Father working	0.93	0.25	0	1	0.07	0.04
Family income:						
quintile 2	0.20	0.40	0	1	0.09	0.02
quintile 3	0.21	0.41	0	1	0.15	0.02
quintile 4	0.21	0.41	0	1	0.23	0.02
quintile 5	0.20	0.40	0	1	0.32	0.03
Mother's schooling:						
lower secondary	0.15	0.35	0	1	0.35	0.02
upper secondary	0.07	0.25	0	1	0.45	0.03
lower tertiary	0.04	0.19	0	1	0.71	0.04
upper tertiary	0.00	0.06	0	1	0.88	0.13
Father's schooling:						
lower secondary	0.14	0.35	0	1	0.31	0.02
upper secondary	0.13	0.34	0	1	0.39	0.02
lower tertiary	0.07	0.26	0	1	0.69	0.03
upper tertiary	0.04	0.19	0	1	0.99	0.04
Born abroad	0.00	0.02	0	1	-0.31	0.27
Father born abroad	0.06	0.24	0	1	-0.08	0.05
Mother born abroad	0.03	0.16	0	1	-0.08	0.04
Østfold	0.06	0.24	0	1	0.03	0.03
Akershus	0.09	0.28	0	1	-0.04	0.03
Hedmark	0.04	0.20	0	1	0.07	0.04
Oppland	0.04	0.20	0	1	0.14	0.04
Buskerud	0.05	0.22	0	1	0.04	0.03
Vestfold	0.05	0.21	0	1	0.09	0.04
Telemark	0.04	0.19	0	1	0.14	0.04
A-Agder	0.02	0.14	0	1	0.24	0.05
V-Agder	0.04	0.19	0	1	0.22	0.04
Rogaland	0.08	0.28	0	1	0.11	0.03
Hordaland	0.11	0.31	0	1	0.20	0.03
Sogn Fj.	0.03	0.17	0	1	0.36	0.04
Møre Roms.	0.07	0.25	0	1	0.26	0.03
S-Tr	0.06	0.24	0	1	0.21	0.03
N-Tr	0.03	0.18	0	1	0.42	0.04
Nordland	0.06	0.24	0	1	0.34	0.03
Troms	0.03	0.18	0	1	0.24	0.04
Finnmark	0.02	0.13	0	1	0.27	0.05

Table 4: **Descriptive statistics for Z_2 with corresponding parameter estimates**

Z2-variables	Mean	St.dev.	Min	Max	Parameter estimate	S.E.
Intercept					7.33	0.04
Manufacturing					0	-
Public services	0.28	0.45	0	1	-0.20	0.03
Private services	0.40	0.49	0	1	-0.04	0.03
Unspecified	0.00	0.04	0	1	-0.23	0.04
General	0.19	0.40	0	1	0.05	0.03
Humanities	0.03	0.18	0	1	-0.10	0.04
Teaching	0.06	0.23	0	1	-0.14	0.05
Technical					0	-
Business/administrative	0.21	0.41	0	1	0.02	0.02
Transport	0.03	0.17	0	1	0.03	0.01
Health	0.05	0.23	0	1	0.07	0.00
Farming/fisheries	0.02	0.13	0	1	-0.06	0.05
Services/military	0.06	0.23	0	1	0.08	0.01
Østfold	0.05	0.21	0	1	-0.15	0.01
Akershus	0.09	0.28	0	1	-0.03	0.01
Oslo					0	-
Hedmark	0.03	0.18	0	1	-0.19	0.00
Oppland	0.03	0.18	0	1	-0.22	0.01
Buskerud	0.05	0.21	0	1	-0.11	0.01
Vestfold	0.04	0.19	0	1	-0.14	0.01
Telemark	0.03	0.17	0	1	-0.13	0.01
A-Agder	0.02	0.13	0	1	-0.18	0.01
V-Agder	0.03	0.17	0	1	-0.14	0.01
Rogaland	0.08	0.27	0	1	-0.02	0.01
Hordaland	0.09	0.29	0	1	-0.11	0.01
Sogn Fj.	0.02	0.15	0	1	-0.18	0.01
Møre Roms.	0.05	0.22	0	1	-0.16	0.01
S-Tr	0.06	0.23	0	1	-0.18	0.01
N-Tr	0.02	0.15	0	1	-0.23	0.01
Nordland	0.05	0.21	0	1	-0.20	0.01
Tromsø	0.03	0.18	0	1	-0.16	0.01
Finnmark	0.02	0.12	0	1	-0.21	0.01

Recent publications in the series Discussion Papers

- 395 T. Harding, H. O. Aa. Solheim and A. Benedictow (2004): House ownership and taxes
- 396 E. Holmøy and B. Strøm (2004): The Social Cost of Government Spending in an Economy with Large Tax Distortions: A CGE Decomposition for Norway
- 397 T. Hægeland, O. Raaum and K.G. Salvanes (2004): Pupil achievement, school resources and family background
- 398 I. Aslaksen, B. Natvig and I. Nordal (2004): Environmental risk and the precautionary principle: "Late lessons from early warnings" applied to genetically modified plants
- 399 J. Møen (2004): When subsidized R&D-firms fail, do they still stimulate growth? Tracing knowledge by following employees across firms
- 400 B. Halvorsen and Runa Nesbakken (2004): Accounting for differences in choice opportunities in analyses of energy expenditure data
- 401 T.J. Klette and A. Raknerud (2004): Heterogeneity, productivity and selection: An empirical study of Norwegian manufacturing firms
- 402 R. Aaberge (2005): Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality
- 403 F.R. Aune, S. Kverndokk, L. Lindholt and K.E. Rosendahl (2005): Profitability of different instruments in international climate policies
- 404 Z. Jia (2005): Labor Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure
- 405 Z. Jia (2005): Retirement Behavior of Working Couples in Norway. A Dynamic Programming Approach
- 406 Z. Jia (2005): Spousal Influence on Early Retirement Behavior
- 407 P. Frenger (2005): The elasticity of substitution of superlative price indices
- 408 M. Mogstad, A. Langørgen and R. Aaberge (2005): Region-specific versus Country-specific Poverty Lines in Analysis of Poverty
- 409 J.K. Dagsvik (2005) Choice under Uncertainty and Bounded Rationality
- 410 T. Fæhn, A.G. Gómez-Plana and S. Kverndokk (2005): Can a carbon permit system reduce Spanish unemployment?
- 411 J. Larsson and K. Telle (2005): Consequences of the IPPC-directive's BAT requirements for abatement costs and emissions
- 412 R. Aaberge, S. Bjerre and K. Doksum (2005): Modeling Concentration and Dispersion in Multiple Regression
- 413 E. Holmøy and K.M. Heide (2005): Is Norway immune to Dutch Disease? CGE Estimates of Sustainable Wage Growth and De-industrialisation
- 414 K.R. Wangen (2005): An Expenditure Based Estimate of Britain's Black Economy Revisited
- 415 A. Mathiassen (2005): A Statistical Model for Simple, Fast and Reliable Measurement of Poverty
- 416 F.R. Aune, S. Glomsrød, L. Lindholt and K.E. Rosendahl: Are high oil prices profitable for OPEC in the long run?
- 417 D. Fredriksen, K.M. Heide, E. Holmøy and I.F. Solli (2005): Macroeconomic effects of proposed pension reforms in Norway
- 418 D. Fredriksen and N.M. Stølen (2005): Effects of demographic development, labour supply and pension reforms on the future pension burden
- 419 A. Alstadsæter, A-S. Kolm and B. Larsen (2005): Tax Effects on Unemployment and the Choice of Educational Type
- 420 E. Biørn (2005): Constructing Panel Data Estimators by Aggregation: A General Moment Estimator and a Suggested Synthesis
- 421 J. Bjørnstad (2005): Non-Bayesian Multiple Imputation
- 422 H. Hungnes (2005): Identifying Structural Breaks in Cointegrated VAR Models
- 423 H. C. Bjørnland and H. Hungnes (2005): The commodity currency puzzle
- 424 F. Carlsen, B. Langset and J. Rattsø (2005): The relationship between firm mobility and tax level: Empirical evidence of fiscal competition between local governments
- 425 T. Harding and J. Rattsø (2005): The barrier model of productivity growth: South Africa
- 426 E. Holmøy (2005): The Anatomy of Electricity Demand: A CGE Decomposition for Norway
- 427 T.K.M. Beatty, E. Røed Larsen and D.E. Sommervoll (2005): Measuring the Price of Housing Consumption for Owners in the CPI
- 428 E. Røed Larsen (2005): Distributional Effects of Environmental Taxes on Transportation: Evidence from Engel Curves in the United States
- 429 P. Boug, Å. Cappelen and T. Eika (2005): Exchange Rate Pass-through in a Small Open Economy: The Importance of the Distribution Sector
- 430 K. Gabrielsen, T. Bye and F.R. Aune (2005): Climate change- lower electricity prices and increasing demand. An application to the Nordic Countries
- 431 J.K. Dagsvik, S. Strøm and Z. Jia: Utility of Income as a Random Function: Behavioral Characterization and Empirical Evidence
- 432 G.H. Bjertnæs (2005): Avoiding Adverse Employment Effects from Energy Taxation: What does it cost?
- 433 T. Bye and E. Hope (2005): Deregulation of electricity markets—The Norwegian experience
- 434 P.J. Lambert and T.O. Thoresen (2005): Base independence in the analysis of tax policy effects: with an application to Norway 1992-2004
- 435 M. Rege, K. Telle and M. Votruba (2005): The Effect of Plant Downsizing on Disability Pension Utilization
- 436 J. Hovi and B. Holtmark (2005): Cap-and-Trade or Carbon Taxes? The Effects of Non-Compliance and the Feasibility of Enforcement
- 437 R. Aaberge, S. Bjerre and K. Doksum (2005): Decomposition of Rank-Dependent Measures of Inequality by Subgroups
- 438 B. Holtmark (2005): Global per capita CO₂ emissions - stable in the long run?
- 439 E. Halvorsen and T.O. Thoresen (2005): The relationship between altruism and equal sharing. Evidence from inter vivos transfer behavior
- 440 L-C. Zhang and I. Thomsen (2005): A prediction approach to sampling design

- 441 Ø.A. Nilsen, A. Raknerud, M. Rybalka and T. Skjerpen (2005): Lumpy Investments, Factor Adjustments and Productivity
- 442 R. Golombek and A. Raknerud (2005): Exit Dynamics with Adjustment Costs
- 443 G. Liu, T. Skjerpen, A. Rygh Swensen and K. Telle (2006): Unit Roots, Polynomial Transformations and the Environmental Kuznets Curve
- 444 G. Liu (2006): A Behavioral Model of Work-trip Mode Choice in Shanghai
- 445 E. Lund Sagen and M. Tsygankova (2006): Russian Natural Gas Exports to Europe. Effects of Russian gas market reforms and the rising market power of Gazprom
- 446 T. Ericson (2006): Households' self-selection of a dynamic electricity tariff
- 447 G. Liu (2006): A causality analysis on GDP and air emissions in Norway
- 448 M. Greker and K.E. Rosendahl (2006): Strategic Climate Policy in Small, Open Economies
- 449 R. Aaberge, U. Colombino and T. Wennemo (2006): Evaluating Alternative Representation of the Choice Sets in Models of Labour Supply
- 450 T. Kornstad and T.O. Thoresen (2006): Effects of Family Policy Reforms in Norway. Results from a Joint Labor Supply and Child Care Choice Microsimulation Analysis
- 451 P. Frenger (2006): The substitution bias of the consumer price index
- 452 B. Halvorsen (2006): When can micro properties be used to predict aggregate demand?
- 453 J.K. Dagsvik, T. Kornstad and T. Skjerpen (2006): Analysis of the discouraged worker phenomenon. Evidence from micro data
- 454 G. Liu (2006): On Nash equilibrium in prices in an oligopolistic market with demand characterized by a nested multinomial logit model and multiproduct firm as nest
- 455 F. Schroyen and J. Aasness (2006): Marginal indirect tax reform analysis with merit good arguments and environmental concerns: Norway, 1999
- 456 L-C Zhang (2006): On some common practices of systematic sampling
- 457 Å. Cappelen (2006): Differences in Learning and Inequality
- 458 T. Borgersen, D.E. Sommervoll and T. Wennemo (2006): Endogenous Housing Market Cycles
- 459 G.H. Bjertnæs (2006): Income Taxation, Tuition Subsidies, and Choice of Occupation
- 460 P. Boug, Å. Cappelen and A.R. Swensen (2006): The New Keynesian Phillips Curve for a Small Open Economy
- 461 T. Ericson (2006): Time-differentiated pricing and direct load control of residential electricity consumption
- 462 T. Bye, E. Holmøy and K. M. Heide (2006): Removing policy based comparative advantage for energy intensive production. Necessary adjustments of the real exchange rate and industry structure
- 463 R. Bjørnstad and R. Nymoen (2006): Will it float? The New Keynesian Phillips curve tested on OECD panel data
- 464 K.M.Heide, E. Holmøy, I. F. Solli and B. Strøm (2006): A welfare state funded by nature and OPEC. A guided tour on Norway's path from an exceptionally impressive to an exceptionally strained fiscal position
- 465 J.K. Dagsvik (2006): Axiomatization of Stochastic Models for Choice under Uncertainty
- 466 S. Hol (2006): The influence of the business cycle on bankruptcy probability
- 467 E. Røed Larsen and D.E. Sommervoll (2006): The Impact on Rent from Tenant and Landlord Characteristics and Interaction
- 468 Suzan Hol and Nico van der Wijst (2006): The financing structure of non-listed firms
- 469 Suzan Hol (2006): Determinants of long-term interest rates in the Scandinavian countries
- 470 R. Bjørnstad and K. Øren Kalstad (2006): Increased Price Markup from Union Coordination - OECD Panel Evidence.
- 471 E. Holmøy (2006): Real appreciation as an automatic channel for redistribution of increased government non-tax revenue.
- 472 T. Bye, A. Bruvoll and F.R. Aune (2006): The importance of volatility in inflow in a deregulated hydro-dominated power market.
- 473 T. Bye, A. Bruvoll and J. Larsson (2006): Capacity utilization in a generalized Malmquist index including environmental factors: A decomposition analysis
- 474 A. Alstadsæter (2006): The Achilles Heel of the Dual Income Tax. The Norwegian Case
- 475 R. Aaberge and U. Colombino (2006): Designing Optimal Taxes with a Microeconomic Model of Household Labour Supply
- 476 I. Aslaksen and A.I. Myhr (2006): "The worth of a wildflower": Precautionary perspectives on the environmental risk of GMOs
- 477 T. Fæhn and A. Bruvoll (2006): Richer and cleaner - at others' expense?
- 478 K.H. Alfsen and M. Greker (2006): From natural resources and environmental accounting to construction of indicators for sustainable development
- 479 T. Ericson (2006): Direct load control of residential water heaters
- 480 J.K. Dagsvik and G. Liu (2006): A Framework for Analyzing Rank Ordered Panel Data with Application to Automobile Demand
- 481 J.K. Dagsvik and Z. Jia (2006): Labor Supply as a Choice among Latent Job Opportunities. A Practical Empirical Approach
- 482 T.A. Galloway (2006): Do Immigrants Integrate Out of Poverty in Norway?
- 483 T.A. Galloway (2006): The Labor Market Integration of Immigrant Men and Women
- 484 J.H. Fiva and M. Rønning (2006): The Incentive Effects of Property Taxation: Evidence from Norwegian School Districts
- 485 M. Greker and Y. Chen (2006): Can voluntary product-labeling replace trade bans in the case of GMOs?
- 486 J.K. Dagsvik, T. Hægeland and A. Raknerud (2006): Estimation of earnings- and schooling choice relations: A likelihood approach
- 487 E. Holmøy (2006): Can welfare states outgrow their fiscal sustainability problems?
- 488 J.K. Dagsvik, M. Locatelli and S. Strøm (2006): Simulating labor supply behavior when workers have preferences for job opportunities and face nonlinear budget constraints