MEMORANDUM

No 11/2005

Tax evasion and labour supply in Norway in 2003:
Structural models versus flexible functional form models

Kari Due-Andresen

Department of Economics
University of Oslo

Tax evasion and labour supply in Norway in 2003:
Structural models versus flexible functional form models

Kari Due-Andresen

Department of Economics, University of Oslo

May 2005

Preface

Counsel has been given by Professor Steinar Strøm, who also initiated this project to compare different functional forms for analysing individual labour supply when tax evasion is an option. The project is financed in part by the Frisch Centre for Economic Research.

Table of contents

1. *Summary*

Current quantitative economic research often suffers from the lack of theoretical principles on which assumptions about functional form can be made. The standard approach in empirical analyses is to let the data determine functional forms within ad hoc selected parametric classes including so-called flexible functional forms. However, this approach is thought unsatisfactory in the context of preference modelling (Dagsvik and Strøm, 2005). The purpose of my analysis is to estimate a model for individual labour supply when tax evasion is an option. I put two different utility model approaches up against each other in an attempt to best approximate the individual's decision making process based on preferences. The first approach I attempt is based on a structural Box-Cox utility model for preferences, while the other approach is based on a flexible utility function in which the preferences are approximated with a polynomial. The models and empirical analysis are based on micro data from an MMI survey carried out in 2003. To run the calculations I have used the data processing software TSP 4.0 and GiveWin.

I have based my analysis on the same logit model for labour supply as is presented in the paper "Labor supply when tax evasion is an option" by Jørgensen, Ognedal and Strøm (2005) where the individual is assumed to make relevant decisions in two stages. The individual first decides whether to work only in the white sector, to work both in the white and the black sector, or only in the black sector. I assume that the individual chooses the strategy yielding the highest utility. To an observer the preferences of the individuals are concealed and therefore preferences are seen as random. Therefore, the probability that an individual will pursue a tax evading strategy or an honest strategy must be estimated. The random elements in the utilities are assumed to be extreme-value distributed. Given the strategy of evading taxes or not, the individual then in the second stage decides the number of hours he or she will work.

In my calculations I have used the probability of being caught for tax evasion based on the respondents' perception of detection probabilities. A probability weighing function, f(q) is included to allow for the possibility that individuals give overweight to small probabilities, see (Kahneman and Tversky, 1979). The specification implies a rank dependent expected utility model, with the expected utility model as a special case; see Quiggin (1982, 1993).

The estimates of the structural models imply that the individual utility function is quasi-concave. This implies that the indifference curves have the "traditional" convex shapes and show the reasonable pattern that the needed compensation for working one hour more, in terms of consumption is sharply increasing with hours of work, equivalent to fewer hours of leisure, for low levels of leisure. The flexible models, on the other hand, produce indifference curves with unreasonable properties given standard assumptions regarding the individual's utility maximization.

All models produce values for labour supply elasticity suggesting that on an aggregated level, comprising both honest and tax evading individuals, wage increase has a positive impact on hours worked in the regular economy and a negative impact on hours supplied in the black economy. However, the models vary in their estimated response to wage increase when looking at labour supply conditional on the chosen tax strategy. The flexible model 3 and the structural model 4 predict increased white hours and reduced black hours, while the flexible models 1 and 2 predict increased white hours but also increased black hours.

Considering the model estimates, predictions and goodness of fit values it seems that the structural Box-Cox utility model I have worked with produce results that are well in line with established assumptions regarding the individual's utility maximization and supply of labour. The flexible models, on the other hand, seem to have an obvious disadvantage for predicting individual utility maximising behaviour when it comes to choosing labour hours, determining leisure and consumption in an environment with a black labour market.

2. Introduction

I will look at how individuals supply labour based on the valuation of income from work in the black and white sectors of the Norwegian economy.

The background for my analysis are the still unsolved problems in economic modelling related to functional form and the distribution of unobservable variables regarding the individual preferences and utility structure. The tradition in economics has been either to employ ad hoc assumption with regard to functional form and the distribution of unobservable variables, or to rely on non-parametric approaches (Dagsvik and Strøm, 2005).

I will attempt to compare two suggested models, one structural model and one flexible model, for individual utility of labour income and the subsequent decision to supply labour when working in the black market is an option. My hypothesis is that the structural model might be better fitting individual utility maximizing behaviour in this setting than the flexible model. The models and empirical analysis are based on micro data from an MMI survey carried out in 2003. To run the calculations I have used the data processing programs TSP 4.0 and GiveWin.

3. Models

I have based my analysis on the same logit model for labour supply as is presented in the paper "Labor supply when tax evasion is an option" by Øystein Jørgensen, Tone Ognedal and Steinar Strøm (2005).

The purpose of my analysis is to estimate a model for individual labour supply when tax evasion is an option. Two different utility model approaches are compared in an attempt to best approximate the individual's decision making process based on preferences. The first approach I attempt is based on a structural model for preferences, while the other approach is based on a flexible utility function. I will draw my conclusions regarding which model is best suited based on how well the model estimates and predicts tax evasion choices and labour supply, as well as how the models comply with the predictions from economic theory.

The individual is assumed to make relevant decisions in two stages. The individual first decides whether to work only in the white sector, in which case I label the respondent "honest" (H), or to work both in the white and black sector, or only in the black sector and therefore be a tax evader (E). I assume that the individual chooses the strategy yielding the highest utility. To an observer the preferences of the individuals are concealed and therefore preferences are seen as random. Therefore, the probability that an individual will pursue a tax evading strategy or an honest strategy must be estimated. The random elements in the utilities are assumed to be extreme-value distributed. The expected utility maximum, across the alternatives in the choice sets, for the two different strategies can be found in closed form. The probabilities of choosing an honest or evading strategy then depend on these expected values of maximum utility. I further assume that the probability of choosing an evading strategy also depends on the individual's perception of how socially acceptable tax evading is and on the opportunities for the individual to evade taxes. Opportunities for tax evasion may vary greatly across employment sectors and occupations.

If the tax evasion is detected by the authorities, the evaders then have to pay a fine in addition to a tax on total income. I will assume that the probability of getting caught is unknown to the individual, who therefore makes his or her tax evasion decision under uncertainty. In accordance

with Allingham and Sandmo (1972) I let the individuals choose their tax evading activities according to expected utility, but I also allow for the possibility that the individual deviates from the expected utility strategy by giving overweight to small probabilities, see Kahneman and Tversky (1979).

In the second stage, given the chosen tax strategy, the individual decides the number of hours he or she will work. With a tax evasion strategy, part of the wage income is not declared to the authorities. Thus, an individual who chooses a strategy of tax evasion may work in the white as well as in the black economy. A person following an honest strategy works only in the white sector of the economy.

The wage rates used to calculate gross earnings equal the hourly wage rate reported by the respondent in the questionnaire. The same wage rate is used in the white as well as in the black economy. Hours worked in the white sector are reported in broad categories and I have used the category midpoints 10, 25, 37.5 hours per week and with 50 hours per week as a maximum. Hours worked in the black economy are reported as annual hours. Again I use the category midpoints, 10, 25, 37, 75, 150, 250 and with 600 as a maximum. This way of treating feasible working hours reflects what is reported by the respondents, namely that working in the black economy has the character of being a side job.

To explain the econometric model I start with analysing the second stage, i.e. hours supplied, given the decision of evading taxes or not.

2nd stage: choosing hours, determining labour supply and leisure

In the second stage the individual, given the chosen strategy of being honest (H) or evade taxes (E), decides on how many hours to work.

I define the following variables:

C_{iH} =   After-tax wage income when the individual follows an honest strategy (H)

$C_{ijE,T}$ =   After-tax and penalty income when the individual is a tax evader (E). The subscript, T, indicates that the tax evasion is detected.

$C_{ijE,NT} =$	As above, but now the tax evasion is not detected, as indicated by the subscript NT
$h_{iH} =$	Annual hours in the white economy; $i = 1,2,..,n$, where n is the number of hour categories. When $i=1$, $h=0$ and the individual does not work.
$h_{ij} = h_{iH}+h_{jE} =$	Annual hours, where h_{iH} is hours worked in the "white" economy and h_{jE} is hours worked in the "black" economy.
$W_H =$	Hourly wage rate in the white economy
$W_E =$	Hourly wage rate in the black economy
$R_{iH} =$	Gross annual wage income $= W_H * h_{iH}$
$R_{jE} =$	Gross annual wage income $= W_E * h_{jE}$
$\tau =$	The fine that the evader has to pay if detected.
$I =$	Non-wage income
$T(R,I) =$	Taxes paid, represented by a step-wise linear function of wage income and non-wage income.
$X =$	A vector of socio-demographic characteristics.
$q =$	The probability of detection ($1 \geq q \geq 0$), as perceived by the respondents
$f(q) =$	The probability weighting function with respect to the perceived likelihood of being caught for tax evasion.
$\varepsilon_{iH} =$	A random variable, assumed to be extreme value iid distributed with zero mean and a constant variance.
$U_{iH} =$	The random utility function when the individual follows an honest strategy and works h_{iH} hours
$U_{ijE} =$	The random utility function when the individual follows a tax evading strategy
$S_H =$	The expected value of the maximum of the utility function across the alternatives in the choice set, given an honest strategy.
$S_E =$	The expected value of the maximum of the utility function across alternatives in the choice set.

Honest strategy

(1) $\quad C_{iH} = R_{iH} + I - T(R_{iH}, I) \; ; i = 1, 2, , , n$

(2) $\quad U_{iH} = u(C_{iH}, h_{iH}, X) + \varepsilon_i \; ; i = 1, 2, , , n$

Here u(.) is the deterministic part of the utility function and ε_i is the random part. The random part may be known to the individual but not to the outside observer. The total utility is in this case an ordinal utility function.

As demonstrated in Ben-Akiva and Lerman (1979) S_H can be represented by:

(3) $\quad S_H = E[max_{i=1,2,,,n} \; U_{iH}] = \mu_2 \ln \sum_{k=1}^{n} \exp(u_{kH} / \mu_2)$

S_H can be interpreted as the consumer surplus associated with the n white alternatives. The constant, μ_2, functions as a scaling coefficient in the deterministic part of the utility function and reflects unobserved heterogeneity in preferences. The larger the value is of μ_2, the more uncertain are the preferences. Naturally, μ_2 is not identified from the data.

The probability of choosing h_{iH} hours, conditional on the honest strategy, is given by:

$P(h_{iH} \mid H) = P(U_{iH} = max_{k=1,2,,,n} \; U_{kH})$

With ε_i being extreme value iid distributed, this optimal choice probability, $P(h_{iH}|H)$, is a multinomial logit. From (3) we observe that this multinomial logit can be derived from taking the derivatives of the consumer surplus, S_H, with respect to the deterministic part of the utility function:

(4) $\quad P(h_{iH} \mid H) = \dfrac{\partial S_H}{\partial u_{iH}} = \dfrac{\exp(u_{iH} / \mu_2)}{\sum_{k=1}^{n} \exp(u_{kH} / \mu_2)} \; ; i = 1, 2, , , n$

Tax evading strategy

(5) $\quad C_{ij,E,T} = R_{iH} + R_{jE} + I - T(R_{iH} + R_{jE}, I) - \tau(R_{jE}); i, j = 1, 2, , , n$

(6) $\quad C_{ij,E,NT} = R_{iH} + R_{jE} + I - T(R_{iH}, I); i, j = 1, 2, , , n$

The probability of being caught for tax evasion is based on the respondents' perception of detection probabilities and is represented by the probability weighing function, $f(q)$. This function is included to allow for the possibility that individuals give overweight to small probabilities (Kahneman and Tversky (1979)). The specification implies a rank dependent expected utility model, with the expected utility model as a special case; see Quiggin (1982, 1993). Detection probabilities, as perceived by the respondents and represented by q, range from 0 to 25%. Actual probability of detection is hard to determine, even for the tax authorities, but putting 25% as the highest detection probability is assumed realistic. The parameter, **a**, is unknown and will be estimated by maximum likelihood based on the responses in the dataset. Then

$$(7) \quad \begin{cases} f(q) = 1 - \dfrac{1}{2}[1 + (1-q)^a - q^a] \text{ for } a \geq 0 \\[2mm] f(q) = \dfrac{1}{2} \text{ for } a = 0 \\[2mm] f(q) = q \text{ for } a = 1 \end{cases}$$

The larger the expression in brackets in the definition of $f(q)$ is, the smaller weight is put on the possibility of being caught, or put differently; the larger is the weight on small perceived probabilities of being caught for tax evasion.

For a>0:

$$(8) \quad \begin{cases} f(1) = 1 \\[1mm] f(0) = 0 \\[1mm] f(\dfrac{1}{2}) = \dfrac{1}{2} \end{cases}$$

Then when a>0:

$$(9) \begin{cases} \text{for all q}: & \dfrac{\partial f(q)}{\partial q} > 0 \\[2ex] \text{for } q = \dfrac{1}{2}: & \dfrac{\partial^2 f(q)}{\partial q^2} = 0 \\[2ex] \text{for } q < \dfrac{1}{2}: & \dfrac{\partial^2 f(q)}{\partial q^2} < 0 \text{ for } 0 < a < 1 \\[2ex] & \dfrac{\partial^2 f(q)}{\partial q^2} > 0 \text{ for } 1 < a < 2 \\[2ex] & \dfrac{\partial^2 f(q)}{\partial q^2} < 0 \text{ for } a \geq 2 \end{cases}$$

The larger the value of q, the larger the value of f(q), as is indicated by the first derivative above In this case, the relevant values of q range between 0 and 0.25. In this range, if the parameter, **a**, takes on a value between 0 and 1, or larger or equal to two, the growth in f(q) as q grows is diminishing, implying that people tend to put larger weight on small possibilities of being detected. This can be seen from the negative sign of the second derivatives in the expression (9) above. The opposite is the case, if the parameter, **a,** takes values between 1 and 2.

$$(10) \quad U_{ijE} = f(q)u(C_{ij,E,T}, h_{iH} + h_{jE}, X) + (1 - f(q))u(C_{ij,E,NT}, h_{iH} + h_{jE}, X) + \varepsilon_{ij}; i, j = 1, 2, , , n$$

This random utility function has two parts, the first part being the deterministic one, describing rank dependent expected utility related to the lottery of tax evasion, and the second part being the random term, following the same distribution as the random term in (2). Expected value of the maximum of the utility function for the tax evader is then:

$$(11) \quad S_E = E[\max_{i=1,2,,n;j=1,2,,n} U_{ijE}] = \mu_2 \ln \sum_{k=1}^{n} \sum_{r=1}^{n} \exp(u_{ijE} / \mu_2)$$

where:

(12) $u_{ijE} = f(q)u(C_{ijE,T}, h_{iH} + h_{jE}, X) + (1 - f(q))u(C_{ijE,NT}, h_{iH} + h_{jE}, X)$

Similar to above the probability of working h_{iH} hours in the "white" economy and h_{iE} hours in the "black" economy, conditional on being a tax evader is:

(13) $P(h_{iH,}h_{jE} \mid E) = \dfrac{\partial S_E}{\partial u_{ijE}} = \dfrac{\exp(u_{ijE}/\mu_2)}{\displaystyle\sum_{k=1}^{n}\sum_{r=1}^{n}\exp(u_{rkE}/\mu_2)}; i, j = 1, 2, , , n$

1st stage: choosing a tax strategy

The individual considers two possible strategies; working in the black sector and evade taxes (E) or being honest (H) and work only in the regular sector, given that hours are determined in an optimal way under either of the two strategies. As outlined in Ben-Akiva (1973) the probability of pursuing a strategy can be evaluated by the expected consumer surpluses. Let P(H) denote the probability of pursuing an honest strategy, and let P(E), equal to 1-P(H), be the probability of choosing the alternative strategy of tax evasion. Then

(14) $P(H) = \dfrac{\exp(S_H/\mu_1)}{\exp(S_H/\mu_1) + \exp(S_E/\mu_1)}$

Using the definitions of S_E and S_H:

(15) $P(H) = \dfrac{\left[\displaystyle\sum_{k=1}^{n}\exp(u_{kH}/\mu_2)\right]^{(\mu_2/\mu_1)}}{\left[\displaystyle\sum_{k=1}^{n}\exp(u_{kH}/\mu_2)\right]^{(\mu_2/\mu_1)} + \left[\displaystyle\sum_{k=1}^{n}\sum_{r=1}^{n}\exp(u_{krE}/\mu_2\right]^{(\mu_2/\mu_1)}}$

μ_1 is a positive constant.

This two-stage modelling of labour supply is a nested multinomial logit model. According to McFadden (1978) the nested multinomial logit model is consistent with the maximization of a random utility function if $\mu_1 \geq \mu_2$. Similar to the function of μ_2 in the second stage of the decision process, the constant μ_1, functions as a scaling coefficient in the deterministic part of the utility function and reflects unobserved heterogeneity in preferences in stage 1 of the decision process. Expression (15) shows that when μ_1 tends to infinity, then P(H) and P(E) tend towards ½, implying that the probability of a person choosing tax evasion or not is "fifty-fifty", and could be represented by a coin toss. Expressions (4) and (13) show that when μ_2 tends to infinity, $P(h_{iH}|H)$ and $P(h_{iH},h_{jE}|E)$ tend towards 1/n, making all combinations of chosen black and/or white labour hours just as likely. Therefore, at extreme values of μ_1 and μ_2, the model degenerates to imply that the individual's choice of how many hours to work and in which sector is made at pure random.

The unconditional probabilities of chosen labour hours in the white and black sectors respectively are given by:

$$(16) \quad P(h_{iH}, H) = P(h_{iH} \mid H)P(H)$$
$$(17) \quad P(h_{iH}, h_{jE}, E) = P(h_{iH}, h_{jE} \mid E)P(E)$$

$P(h_{iH},H)$ is the probability of h_{iH} hours worked in the white sector by the so-called "honest" individuals, and $P(h_{iH},h_{jE},E)$ is the probability of the combination of h_{iH} and h_{jE} hours worked in the white and black sectors respectively by the "dishonest" tax evading individuals.

Let N_H be the honest individuals in the sample. These are the ones who in the survey claim not to have evaded taxes during the last twelve months. Then let N_E be the group of tax evaders in the sample. Let subscript s indicate individual. The joint a priori probability of the observations is then given by the likelihood function, L:

$$(18) \quad L = \prod_{s \in N_H} P_s(h_{iH}, H) \prod_{s \in N_E} P_s(h_{iH}, h_{jE}, E)$$

The joint density is considered a function of the parameters rather than as a function of the observations. The likelihood function gives the probability of observing the data given by the respondents in the dataset as a function of the parameters, which will be specified below, one set for the structural models and another set for the flexible models. The maximum likelihood estimate (mle) of the parameters are those values that maximize the likelihood, that is, makes the reported data most likely to occur (Rice, 1995). Rather than maximizing the likelihood function itself it is easier to maximize its natural logarithm.

To simplify matters I will in my analysis concentrate on the special case when $\mu_2/\mu_1=1$. The nested multinomial logit model then degenerates to a multinomial logit model. From (4), (13) and (15) we then get

$$(19) \quad P(h_{iH}, H) = \frac{\exp(u_{iH}/\mu_2)}{\sum_{k=1}^{n}\exp(u_{kH}/\mu_2) + \sum_{k=1}^{n}\sum_{r=1}^{n}\exp(u_{krE}/\mu_2)}$$

and

$$(20) \quad P(h_{iH}, h_{jE}, E) = \frac{\exp(u_{ijE}/\mu_2)}{\sum_{k=1}^{n}\exp(u_{kH}/\mu_2) + \sum_{k=1}^{n}\sum_{r=1}^{n}\exp(u_{krE}/\mu_2)}$$

Current quantitative economic research often suffers from the lack of theoretical principles on which assumptions about functional form can be made. The standard approach in empirical analyses is to let the data determine functional forms within ad hoc selected parametric classes including so-called flexible functional forms. However, this approach is thought unsatisfactory in the context of preference modelling (Dagsvik and Strøm, 2005). I will proceed to empirically compare versions of a structural utility model with versions of a flexible model in which the preferences are approximated with a polynomial. The purpose is to find whether the structural model better approximates the observed data than the flexible model.

3.1.　　Structural models: model 4 and 5

In this section I analyse a structural utility model and look at different versions based on differing working hour input. Model 4 is based on the whole data set of respondents and on the amount of both white and black working hours chosen by the population of respondents. Model 5 is based only on hours worked in the white sector. I estimate two versions of model 5, one based on the hours worked by respondents who claim to only work in the white sector, called model 5a, and model 5b, based on the hours worked in the white sector by all respondents.

I will assume that the inclination of an individual to evade taxes is stronger the more widespread tax evasion is in the population. To capture this possible influence on the individual's decision I assume that the probability of choosing a tax evading strategy depends on the individual's own perception of how socially acceptable tax evasion is. This perception is measured by the observed variable Z_1. Moreover, the opportunity to evade taxes will necessarily differ across employment sectors and positions at the workplace. I will assume that the Norwegian construction sector and the government sector are at opposite extreme ends of an imagined tax evading opportunities scale, meaning that I expect workers in the construction sector to be able to relatively easy keep income from the tax authorities while this possibility is very small for a worker in the government sector. Two dummy variables are introduced to reflect these possible differences in tax evasion opportunities, one for those working in the construction sector (Z_2) and one for those working in the government sector (Z_3). To bring these aspects into the econometric model I weight the expected utility value of choosing a tax evasion strategy by a social norm and opportunity density function, $g(Z_1, Z_2, Z_3)$ or shorter $g(Z)$.
The function $g(Z)$ is represented as:

$$(21) \quad g(Z) = \exp\left(g_0 + g_1 Z_1 + g_2 Z_2 + g_3 Z_3\right)$$

In the Z-vector Z_1, equals 1 if the respondent thinks tax evasion is generally accepted, and equals zero otherwise. Z_2, equals 1 it the respondent works in the construction sector, and equals zero otherwise. Z_3, equals 1 if the respondent works in the government sector, and equals zero otherwise.

Instead of (14) and (15) I then get

$$(14a) \quad P(H) = \frac{\exp(S_H / \mu_1)}{\exp(S_H / \mu_1) + g(Z)\exp(S_E / \mu_1)}$$

and

$$(15a) \; P(H) = \frac{\left[\sum_{k=1}^{n} \exp(u_{kH} / \mu_2)\right]^{(\mu_2/\mu_1)}}{\left[\sum_{k=1}^{n} \exp(u_{kH} / \mu_2)\right]^{(\mu_2/\mu_1)} + g(Z)\left[\sum_{k=1}^{n}\sum_{r=1}^{n} \exp(u_{krE} / \mu_2)\right]^{(\mu_2/\mu_1)}}$$

The norm and opportunity density $g(Z)$ can be interpreted as a threshold level related to choice probabilities. From (14) I get that $P(E) > P(H)$ if $\dfrac{\exp\left(S_H / \mu_1\right)}{\exp\left(S_E / \mu_1\right)} < g(Z)$ or $\dfrac{\exp\left(S_E / \mu_1\right)}{\exp\left(S_H / \mu_1\right)} > \dfrac{1}{g(Z)}$, that is if $\left[S_E - S_H\right] > -\mu_1 \ln g(Z)$. Thus, the probability of being a tax evader will exceed the probability of being honest if the difference in expected consumer surplus exceeds the threshold level $-\mu_1 \ln g(Z)$. And the larger the value of $g(Z)$, the lower the threshold value, meaning that individuals are more prone to work black hours.

When $\mu_2/\mu_1 = 1$, the multinomial logit model corresponding to (19) and (20) becomes

$$(19a) \quad P(h_{iH}, H) = \frac{\exp(u_{iH} / \mu_2)}{\sum_{k=1}^{n} \exp(u_{kH} / \mu_2) + g(Z)\sum_{k=1}^{n}\sum_{r=1}^{n} \exp(u_{krE} / \mu_2)}$$

and

$$(20a) \quad P(h_{iH}, h_{jE}, E) = \frac{g(Z)\exp(u_{ijE} / \mu_2)}{\sum_{k=1}^{n} \exp(u_{kH} / \mu_2) + g(Z)\sum_{k=1}^{n}\sum_{r=1}^{n} \exp(u_{krE} / \mu_2)}$$

Specifying the utility function:

Let the utility function be specified by $v(C,h,X)=u(C,h,X)/\mu_2$. This deterministic part of the utility function is assumed to be a Box-Cox transformation of disposable income and leisure. A justification for this functional form is given in Dagsvik and Strøm (2005).

$$(21) \quad v(C,h,X) = (\alpha_0)\frac{(C/100000)^\lambda - 1}{\lambda} + (\beta_0 + \beta_1 X_1 + \beta_2 X_2)\frac{(8760 - h)^\gamma - 1}{\gamma}$$

where C is disposable income (in 2003 NOK) defined in (1). The variable C captures all details of the step-wise tax-functions. The expression in brackets, (8760-h), is total annual hours minus reported hours worked and represents annual leisure. The fine included in the tax functions when tax evasion is detected, is based on the perceived fines as reported by the respondents. The probability of detection is based on the respondents' perception of detection probabilities and is represented by the probability weighing function, f(q), presented above. X_1 is a dummy representing age (in years) and X_2 is a dummy, equalling 1 if the individual is a woman and zero otherwise.

In estimating model 4 I base my analysis on the whole data set. Model 5, however, is based only on the reported white hours. Model 5a is based on white hours reported by only the honest individuals, i.e. those who report not to evade taxes, while model 5b is based on the reported white hours of all respondents.

I will estimate the unknown parameters, $\alpha_0, \beta_0, \beta_1, \beta_2, \gamma, \lambda$ and g_0, g_1, g_2, g_3 and **a** using the maximum likelihood method.

3.2.   Flexible models: model 1 to 3

I proceed to estimate three versions of a basic flexible model, approximating the utility function concerning chosen consumption and leisure by a polynomial of degree 3, implying a cubical utility function. The cubical functions are in general complicated to analyse since they drastically

17

change form as the parameter values change. The polynomial is described by nine parameters, denoted β_1 to β_9.

The expected utility structure given by the cubic utility function implies that when one of the goods, consumption or leisure, increases the utility increases. At low levels of a good, the utility should increase more as the amount of the good increases, but less and less as the level of good increases. These features of increasing utility, but decreasing deterministic marginal utility, would be represented by positive first-order derivatives and negative second-order derivatives.

$$v[C,h,X] = \beta_1 C + (\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2)L(h) + \beta_3 C^2 + \beta_4[L(h)]^2 + \beta_5 CL(h) + \beta_6 C^3 + \beta_7[L(h)]^3 + \beta_8 C^2 L(h) + \beta_9 C[L(h)]^2$$

where $L(h) = \dfrac{8760 - h}{8760}$, i.e. leisure as a share of total annual hours

In models 2 and 3 all beta values are included, while in model 1 β_9, describing consumption multiplied by leisure to the power of 2 is left out. The exclusion of β_9 is done only to improve the performance of the model.

Model 1:

$$v[C,h,X] = \beta_1 C + (\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2)L(h) + \beta_3 C^2 + \beta_4[L(h)]^2 + \beta_5 CL(h) + \beta_6 C^3 + \beta_7[L(h)]^3 + \beta_8 C^2 L(h)$$

In model 2 the unknown variable coefficient, **a**, is estimated with maximum likelihood, while in models 1 and 3 the variable, **a**, is set to equal zero. By setting the unknown parameter, **a**, to zero the probability weighing function, f(q), is set to ½ for all values of q, which represents the perceived probability of tax evasion being detected by the authorities. The utility function (10) presented above then becomes:

$$(10')\quad U_{ijE} = \frac{1}{2}u(C_{ij,E,T}, h_{iH} + h_{jE}, X) + \frac{1}{2}u(C_{ij,E,NT}, h_{iH} + h_{jE}, X) + \varepsilon_{ij}; i,j = 1,2,,,n$$

To sum up the models:

Model 1: excluding β_9 and setting **a**=0

Model 2: including β_9 and estimating **a**

Model 3: including β_9 and setting **a**=0

I will estimate the unknown parameters, β_1 to β_9 and **a** using the maximum likelihood method.

4. Data

The data I have used are taken from a survey done by a private Norwegian polling institute MMI in the year 2003. The recruiting of participants was done by MMI over the telephone asking the person in the household, above 15 years of age and who most recently celebrated his or her birthday if he/she wanted to participate in a research study. The recruitment was conducted randomly in the Norwegian population.

The survey gathers information regarding relevant personal characteristics of the respondents, such as age and employment, economic variables such as income and taxes, and people's engagement in as well as attitudes towards non-reported income activities.

The answer percentage is fairly high. 73% of the people asked filled out the questionnaire, implying that 62% of the persons initially contacted ended up participating in the survey. This is very good compared to response rates in other surveys, for instance in the consumer expenditure surveys of Statistics Norway.

A relevant question regarding the results is the one of possible systematic bias. A common experience with surveys is that people agreeing to participate might have better knowledge of and a higher interest in the subject in question than the people refusing to participate. The participants might also have "an agenda" when answering. However, the 2-staged process of recruiting and filling out of questionnaires allows for some control of the possible bias. In addition to drawing the recruiting areas randomly the results have afterwards been weighted as if everyone agreed to participate and filled out the questionnaire.

5. Estimates and predictions

5.1.   The structural models 4 and 5

5.1.1. Estimated utility models

Since the parameters g_0, g_1, g_2, g_3 and **a** are connected to the choice of black labour hours, these are not estimated in the models 5a and 5b, based only on white labour hours. Moreover in the structural models 4, 5a and 5b it turned out that the exponent, γ, was close to 0. This indicates that the following utility function is log-linear in leisure. The coefficient $\mu = \mu_2 / \mu_1$ turned out to be close to 1.

Runing the models gave the following estimation results:

Table 1. The estimates of model 4

```
Number of observations = 626     Log likelihood = -1068.74
        Schwarz B.I.C. = 1100.93

                       Standard
 Parameter  Estimate     Error       t-statistic   P-value
 BETA0      10.5402     3.04810       3.45794       [.001]
 BETA1       .026100     .052726       .495021      [.621]
 BETA2      3.23742     1.12738       2.87164       [.004]
 A           .356681     .202998      1.75707       [.079]
 G0         -4.84014     .288017     -16.8051       [.000]
 G1          .970942     .293171      3.31186       [.001]
 G2         1.20287      .378039      3.18187       [.001]
 G3         -.642699     .752093      -.854548      [.393]
 ALFA       3.01760      .379243      7.95692       [.000]
 LAMBDA      .719574     .076902      9.35701       [.000]
```
Source: TSP 4.0

Table 2. The estimates of model 5a:

```
Number of observations = 626      Log likelihood = -563.698
        Schwarz B.I.C. = 579.796


                        Standard
 Parameter  Estimate     Error       t-statistic   P-value
 BETA0      40.9086      4.27719      9.56436       [.000]
 BETA1      .037923      .062055      .611122       [.541]
 BETA2      4.32418      1.41546      3.05497       [.002]
 ALFA       8.91315      .595115      14.9772       [.000]
 LAMBDA     .719926      .039153      18.3876       [.000]
```

Source: TSP 4.0

Table 3. The estimates of model 5b:

```
Number of observations = 626      Log likelihood = -635.431
        Schwarz B.I.C. = 651.529

                        Standard
 Parameter  Estimate     Error       t-statistic   P-value
 BETA0      36.9288      3.97293      9.29509       [.000]
 BETA1      .034777      .059228      .587177       [.557]
 BETA2      5.02718      1.32026      3.80772       [.000]
 ALFA       8.35552      .543467      15.3745       [.000]
 LAMBDA     .697889      .040620      17.1810       [.000]
```

Source: TSP 4.0

A sufficient condition for the deterministic part of the utility function, v(.) to be quasi-concave function is that both λ and γ are less than 1. When they are both equal to 1, the utility function is linear. When they approach zero, the utility function tends towards a log-linear function of consumption and leisure. As already mentioned γ has been set equal to zero, so that the utility function is log-linear in leisure. All models produce estimated values of γ and λ that are both below 1, which means that the estimated deterministic part of the utility function is quasi-concave. The estimates of λ in the different models are quite similar and they are also in line with the estimate of labour supply when tax evasion is ignored; see for instance in Dagsvik and Strøm (2005). In the paper of Dagsvik and Strøm labour supply is estimated on data where tax evasion is ignored. The estimate is 0.64.

In models 4-5b the impact of age has no significant impact on the marginal utility of leisure, while gender has. The estimates imply that women have a higher utility of leisure than men.

Since a higher value of g(Z) implies a lower threshold level for tax evasion I would expect g_1 and g_2 to be positive and g_3 to be negative. The estimates of the g-function are all statistically significant, as is indicated by the p-values. As expected the estimate of g_1, representing the opinion that tax evasion is generally accepted is positive, is positive. Is seems reasonable, that having the opinion that tax evasion is common will make people more prone for tax evasion. Also as expected, the estimated value of g_2, representing work in the construction sector is positive. As presumed, it seems people employed in the construction sector are more likely to work black hours. This could perhaps be explained by an attitude in this sector that tax evasion is ok and/or it may be due to the fact that it is easier to work black hours in this sector than other sectors without the authorities finding out. The opposite seems to be the case for the government sector, as is implied by the negative estimated value of g_3.

Model 4 estimates the parameter, **a,** at 0.357. Since **a** then lies in the range between 0 and 1, this means that for values of q between 0 and 0.25 the growth in f(q) as q grows is diminishing, implying that people tend to put an overweight on small possibilities of being detected. However, as the p-value shows one cannot reject the hypothesis that the parameter, **a,** is equal to zero, meaning that **a** is not significantly different from 0. As described above, if a=0, f(q)=1/2, meaning that the individuals perceive the probability of being caught for tax evasion as "fifty-fifty".

Overall the structural models produce results that are well in line with established assumptions regarding the individual's utility maximization and supply of labour.

5.1.2. Goodness of fit: log likelihood values and McFadden's rho (ρ^2)

To analyse goodness of fit I introduce the commonly used measure of McFadden's rho-squared. McFadden's rho is calculated by logistic regression, and is conceptually similar to an R-squared in linear regression. This measure is defined as (McFadden, 1974): $\rho^2 = 1 - \left[L\left(\beta^*\right) / L\left(0\right) \right]$

23

where, $L(\beta^*)$ is the log likelihood with estimated optimal parameters and $L(0)$ is the log likelihood when all parameters are set to zero, meaning that all combinations of supplied labour hours, white and black, have an equal chance of being chosen.
http://water.usgs.gov/pubs/of/2003/ofr03500/pdf/ofr03-500.pdf
http://www.bwk.tue.nl/urb/p/Danielle/acsp99.html

Table 4: McFadden's Rho-squared for the structural models

McFadden's Rho:		
Mod4	Mod5a	Mod5b
0,552	0,441	0,369

For model 4 the goodness of estimate is 0.552, while it for the models 5a and 5b is 0.441 and 0.369 respectively. This means that all the structural models produce satisfying goodness of fit values. Compared to a model where all choices are made at pure random the structural models explains the observed data 55.2, percent 44.1 percent and 36.9 percent better respectively.

5.2.   The flexible models 1 to 3

5.2.1. Estimated utility models

The flexible models estimated these utility function structures:

Table 5. The estimate of model 1:

```
Number of observations = 626      Log likelihood = -1510.35
        Schwarz B.I.C. = 1542.54

                        Standard
 Parameter  Estimate     Error       t-statistic   P-value
 BETA20     -519.731     66.1799      -7.85330      [.000]
 BETA21     .073214      .063271      1.15715       [.247]
 BETA22     6.90358      1.39233      4.95830       [.000]
 BETA1      98.5632      15.5674      6.33139       [.000]
 BETA3      -21.4468     3.98711      -5.37903      [.000]
 BETA4      1011.78      140.142      7.21973       [.000]
 BETA5      -93.0256     16.1736      -5.75170      [.000]
 BETA6      .811276      .229349      3.53730       [.000]
 BETA7      -496.095     74.1220      -6.69296      [.000]
 BETA8      18.2692      3.24474      5.63043       [.000]


Source: TSP 4.0
```

Table 6. The estimate of model 2:

```
Number of observations = 626      Log likelihood = -1364.52
        Schwarz B.I.C. = 1403.15

                        Standard
 Parameter  Estimate     Error       t-statistic   P-value
 BETA20     377.255      57.5126      6.55952       [.000]
 BETA21     .065960      .063891      1.03238       [.302]
 BETA22     6.68932      1.43027      4.67696       [.000]
 A          -.222780     .012124      -18.3749      [.000]
 BETA1      -214.155     30.6092      -6.99641      [.000]
 BETA3      24.9328      5.14212      4.84874       [.000]
 BETA4      -880.744     128.773      -6.83950      [.000]
 BETA5      432.152      59.3539      7.28094       [.000]
 BETA6      -.809979     .240661      -3.36565      [.001]
 BETA7      500.569      71.6313      6.98813       [.000]
 BETA8      -26.2181     5.15380      -5.08715      [.000]
 BETA9      -204.870     28.5340      -7.17985      [.000]


Source: TSP 4.0
```

Table 7. The estimate of model 3:

```
Number of observations = 626      Log likelihood = -1441.06
        Schwarz B.I.C. = 1476.47

                         Standard
 Parameter  Estimate      Error       t-statistic   P-value
 BETA20     205.884      103.966       1.98029      [.048]
 BETA21     .077761      .062670       1.24079      [.215]
 BETA22     6.99280      1.42379       4.91141      [.000]
 BETA1      -247.351     42.8700       -5.76980     [.000]
 BETA3      27.1618      7.68148       3.53602      [.000]
 BETA4      -591.041     223.487       -2.64463     [.008]
 BETA5      569.389      75.2445       7.56718      [.000]
 BETA6      -.698202     .358204       -1.94917     [.051]
 BETA7      381.914      119.788       3.18824      [.001]
 BETA8      -33.8451     7.31538       -4.62657     [.000]
 BETA9      -302.619     32.4654       -9.32128     [.000]
```

Source: TSP 4.0

Since the models suggested only consider static utility, uncertainty regarding future events is not taken in as an aspect for analysis. Individual utility is analysed as if a person only lives for one period. The individual chooses consumption and hours of labour to supply. Empirical evidence, including the MMI survey this thesis is built upon, suggests that most households have little wealth. Their consumption therefore tracks their income quite closely. As a result their current labour income has a large role in the determination of their consumption. In general, an individual therefore cannot choose its leisure and consumption independently. If an individual increases its labour supply by a small number of hours and uses the resulting extra income to increase consumption, it would be reasonable to expect that utility remains unchanged, meaning that increased consumption weighs up for reduced leisure. Given standard assumptions regarding individual utility, I would expect the first derivative of the deterministic utility with respect to consumption and leisure to be positive also at zero consumption and leisure, meaning that an increase in either consumption or leisure will increase utility for the individual. I would therefore expect the parameters β_1 and $\beta_2 = (\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2)$ describing the linear variables for consumption and leisure to be positive.

In what follows I will focus my discussion on the deterministic part of the utility function, but it should be mentioned that the random part of the utility function may depend on consumption and

leisure. Hence, it is not straightforward to derive marginal utility of consumption, leisure and indifference curves. With these remarks in mind I proceed with a discussion of the implications of model 1-3 on marginal utilities and indifference curves.

As expected model 1 estimate a positive value for β_1 indicating increased utility by increased consumption. Contrary to what is reasonable the model estimates a negative value for $\beta_2 = (\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2)$, describing leisure, implying that more free time, given that initial leisure is zero, should lower utility. The models 2 and 3 also estimate opposite signs for consumption and leisure, but negative for β_1 and positive for $\beta_2 = (\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2)$. Marginal utility should normally decline as the level of either of the goods increases. This feature would manifest itself through negative second derivatives. Then for consumption it should be that:

$$\frac{\partial v}{\partial C} = \beta_1 + 2\beta_3 C + \beta_5 L + 3\beta_6 C^2 + 2\beta_8 LC + \beta_9 L^2 > 0$$

$$\frac{\partial^2 v}{\partial C^2} = 2\beta_3 + 6\beta_6 C + 2\beta_8 L < 0$$

However, given known household behaviour, one might expect that the level of consumption has implications for savings behaviour and thereby labour supply. A third derivative equal to zero, which could indicate that deterministic utility is quadratic rather than cubic, would imply that marginal utility reaches zero at some finite level of consumption and then becomes negative. It might be more reasonable to expect that marginal utility decreases, but that the rate of reduction becomes less and less negative so that the marginal utility stays positive for finite levels of consumption. This would be implied by a positive third derivative. I would therefore expect the third derivative of utility with respect to consumption to be zero or positive (Romer, 2001). Then

$$\frac{\partial^3 v}{\partial C^3} = 6\beta_6 \geq 0$$

As can be seen from the p-values, β_{21}, describing age of the individual, is not significantly different from zero. All the other variables have p-values suggesting they are statistically significant.

If my hypothesis of the individual's valuation of consumption level is correct, the estimated β_6-value should be non-negative. Model 1 estimate a positive, but very low value for β_6. The models 2 and 3 estimate negative values for β_6, implying that marginal utility falls faster and faster as the level of consumption increases. This does not seem plausible.

Already after analysing the first parameters of the models, it seems these flexible models are not able to reproduce expected utility maximizing behaviour.

Model 2 is the only model that estimates a value for the probability weighting parameter, **a**. The estimated value is -0.22. Since the probability weighting function is only defined for non-negative values of **a**, this further suggests that model 2 is not an optimal model.

Considering the parameter values estimated by models 1 to 3, the flexible models seem to have an obvious disadvantage when it comes to predicting individual utility maximising behaviour when it comes to choosing labour hours, determining leisure and consumption in a setting where tax evasion is an option.

5.2.2. Goodness of fit: log likelihood values and McFadden's rho (ρ^2)

The flexible models produce weaker goodness of fit values than the structural model 4, which is the relevant model for comparison, since the models 5a and 5b are only based on white hours.

According to the estimated values for McFadden's rho-squared presented in table 8 below, model 1, model 2 and model 3 explain the observed data respectively 36.6 percent, 42.7 percent and 39.5 percent better than a model where all choices are made at pure random.

Table 8. McFadden's Rho-squared-values for the flexible models.

McFadden's Rho:		
Mod1	**Mod2**	**Mod3**
0,366	0,427	0,395

5.3. Labour supply elasticity

I want to look at how the labour supply of black and white hours is affected by a change in hourly wage. The elasticity is obtained by calculating the relative change in labour supply over all individuals in the sample that results from a one per cent wage increase assuming all else is held constant.

In general real wage appears to be only moderately pro-cyclical. Empirically, there is little support for hypothesis of highly elastic labour supply; rather it seems to be relatively inelastic. In large, this picture is supported largely by the all models I have estimated.

Table 9. Wage elasticity.

Variables	Mod 1 Ex beta9_a=0	Mod 2 Incl beta9	Mod 3 Incl beta9_a=0	Mod 4	Mod 5a White-honest	Mod 5b White-all
El_P(H)	0,191	0,698	0,421	0,029		
El_P(E)	-0,542	-2,055	-0,493	-0,298		
El_(L_H|H)	0,206	0,130	0,139	0,253	0,455	0,410
El_(L_H|E)	0,745	0,194	0,785	0,068		
El_(L_E|E)	0,500	0,172	-0,424	-0,126		
El_L_H	0,473	0,323	0,524	0,235		
El_L_E	-0,055	-1,882	-1,019	-0,410		

All models produce a positive value for El_L_H and a negative value for El_L_E, suggesting that on an aggregated level, comprising both honest and tax evading individuals, wage increase has a

positive impact on hours worked in the regular economy and a negative impact on hours supplied in the black economy. This observation seems reasonable assuming individuals want to "do the right thing" and since increased wages also implies increased fines if tax evasion is detected, meaning that when hourly wages increase, working white hours becomes more attractive than working black hours both morally and economically. Estimated labour supply elasticity ranges from 0.23 to 0.52 for white hours, depending on model and whether the individual chooses to work only in the white sector or in both the black and white sectors.

The models vary in their estimated response to wage increase when looking at labour supply conditional on chosen strategy of tax evasion or not. The models 3 and 4 predict increased white hours and reduced black hours. Models 1 and 2 estimate increased white hours but also increased black hours. Models 5a and 5b, based on the assumption that working black hours is impossible, predict the highest labour supply elasticity of the models at 0.46 and 0.41 respectively. According to these models an honest individual supplies white labour hours more elastically than those who would also choose to work black hours had it been possible. Model 2, which allows for a large weight on small possibilities of getting caught for tax evasion, estimates the lowest labour supply elasticity at 0.13, indicating that hourly wage changes has relatively low impact on their choice of labour supply.

When analysing labour supply elasticity it is important to consider individual level of wage and of wealth as well as amount of black labour supplied. A person with a high wage level and/or considerable wealth might want more leisure if wages are increased and therefore supply fewer hours. In this model framework personal wealth is reported only as return to wealth, that is as "non-wage income" and is included in the consumption expressions of the utility models. Since these models predict increased labour supply when hourly wages increase, it indicates that on an aggregate level, the wealth is not high enough for people to choose more leisure over increased income from increased labour supply when wages go up.

In general it is reasonable to expect that work in the black market for most individuals has the characteristics of a side job. However, should a person for some reason receive a dominant share of his or her income from the black sector, he or she might want to increase the supply of black

labour when wages increase. None of the individuals in the survey report to work more in the black than in the white sector, and I therefore think it reasonable that black labour supply is reduced as it is substituted by increased white labour supply when wages increase. The results from the models 1 and 2 therefore seem unrealistic.

Summing up, the models indicate that elasticity of labour supply is relatively low. However, when work in the white sector is the only option for income, labour supply is considerably more elastic than when black labour income is also possible and even more so when tax evasion is considered relatively safe.

5.4. Comparing the models 4 and 5 with the models 1 to 3

5.4.1. Estimates and indifference curves

As I have already established from the estimated results, the estimates of the structural models imply that the utility function is quasi-concave in the central area of consumption and leisure combinations. This implies that the indifference curves have the "traditional" convex shapes, as I have shown graphically below. The models also show the reasonable pattern that the needed compensation for working one hour more, in terms of consumption is sharply increasing with hours of work, equivalent to fewer hours of leisure, for low levels of leisure.

The indifference curves are estimated for men and women separately and are represented by a smooth line for men and a stippled line for women in the graphs.
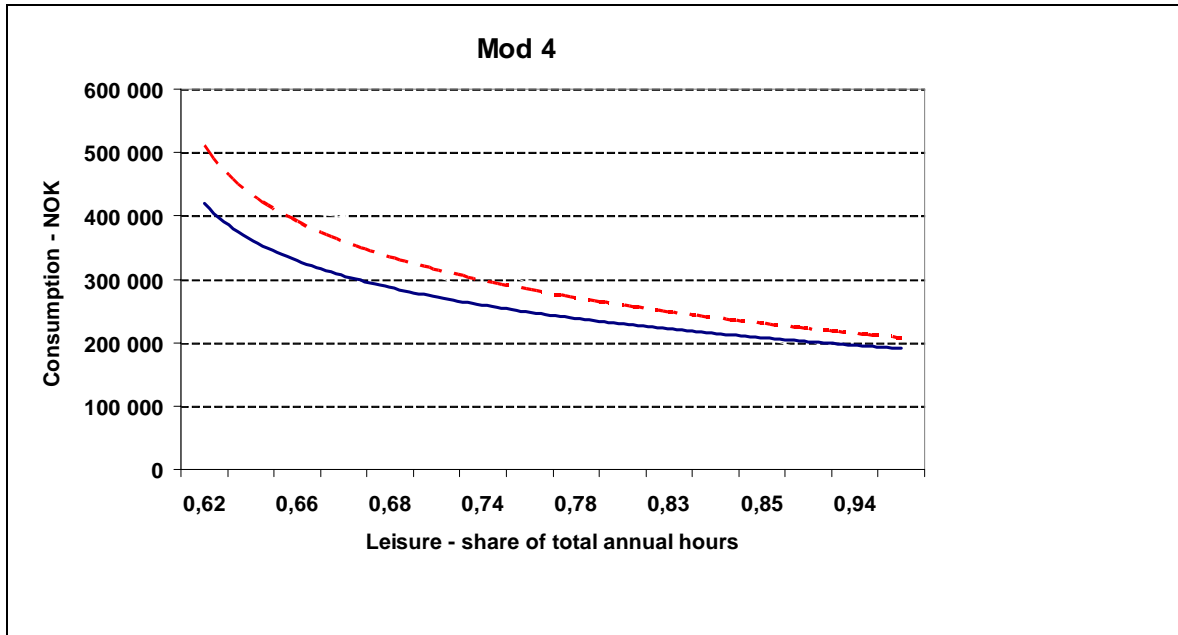
31

Fig. 1. Indifference curves of model 4.

Mod 4

Consumption - NOK (y-axis): 0, 100 000, 200 000, 300 000, 400 000, 500 000, 600 000

Leisure - share of total annual hours (x-axis): 0,62 0,66 0,68 0,74 0,78 0,83 0,85 0,94

Fig. 2. Indifference curves for model 5a.

Mod 5a (White - honest)

Consumption - NOK (y-axis): 0, 100 000, 200 000, 300 000, 400 000, 500 000, 600 000

Leisure - share of total annual hours (x-axis): 0,62 0,66 0,68 0,74 0,78 0,83 0,85 0,94

Fig. 3. Indifference curves for model 5b.

Mod 5b (White - all)

Consumption - NOK vs *Leisure - share of total annual hours*
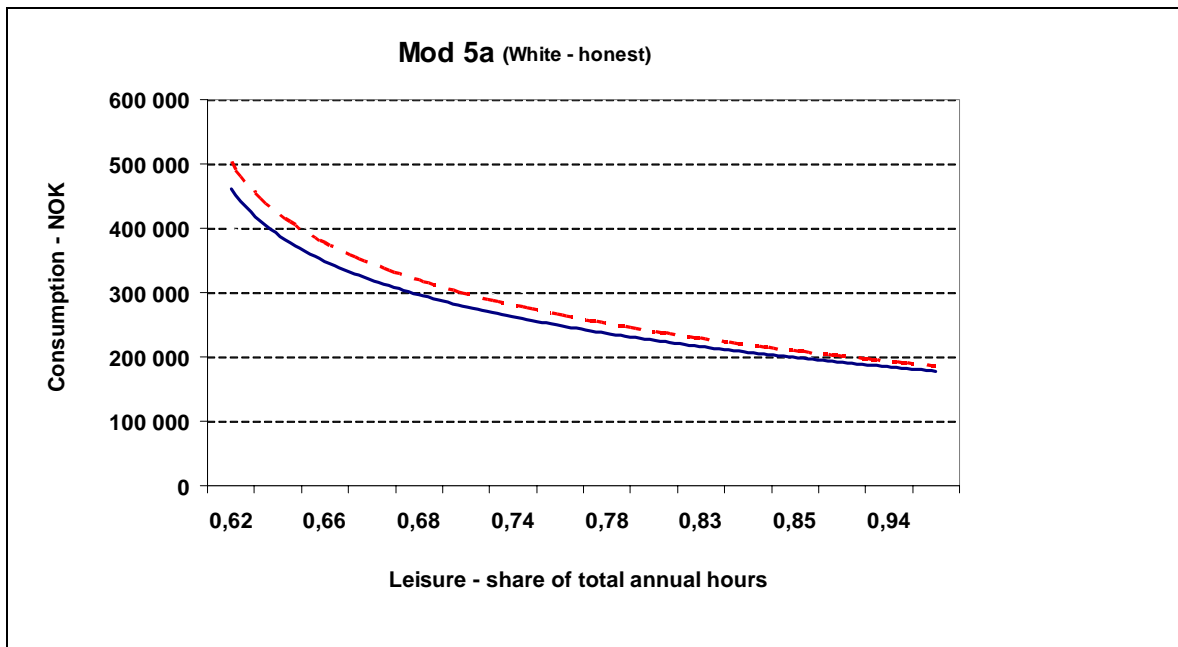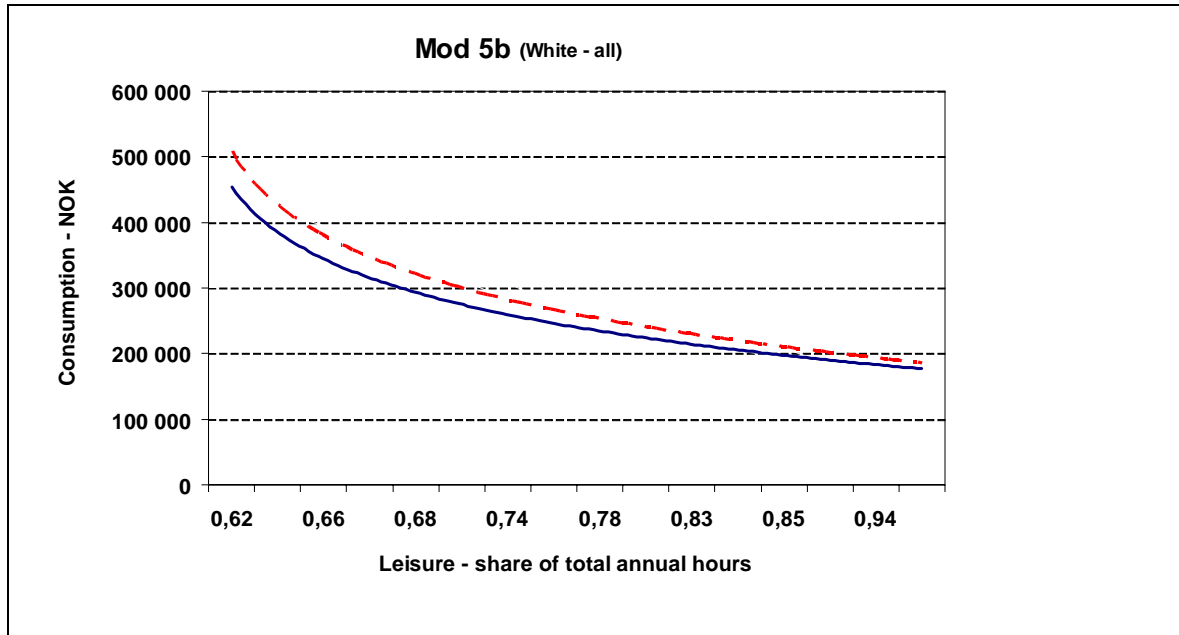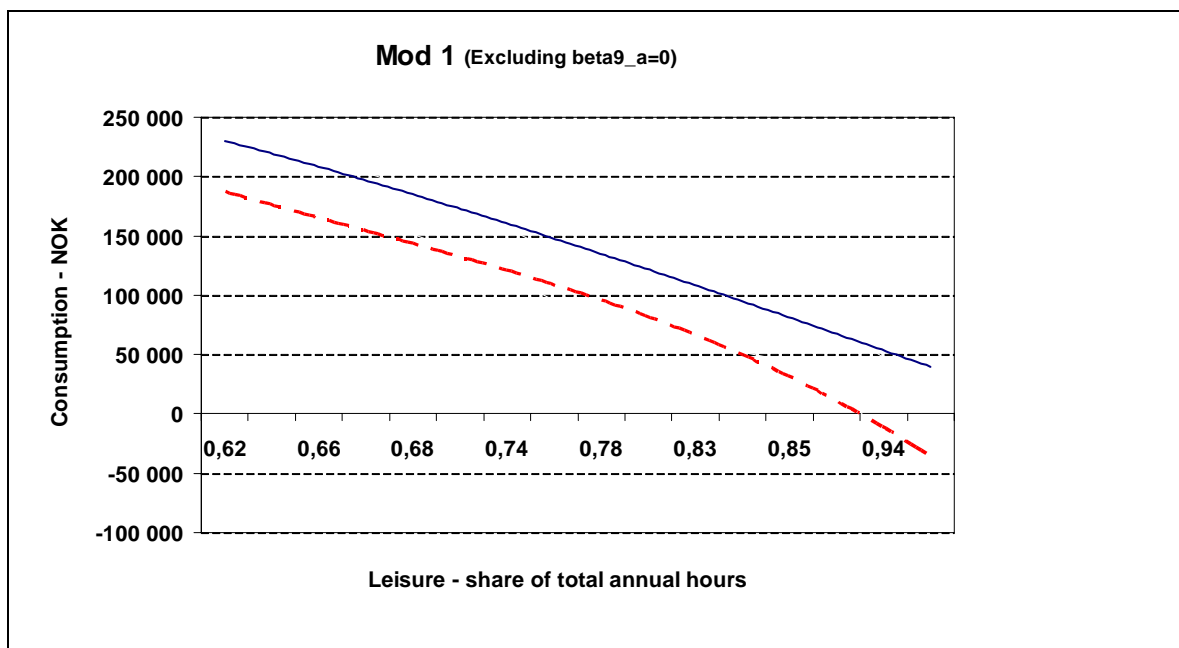
As expected the flexible models 1 to 3, produce indifference curves with unreasonable properties given standard assumptions regarding the individual's consumption and leisure preferences.
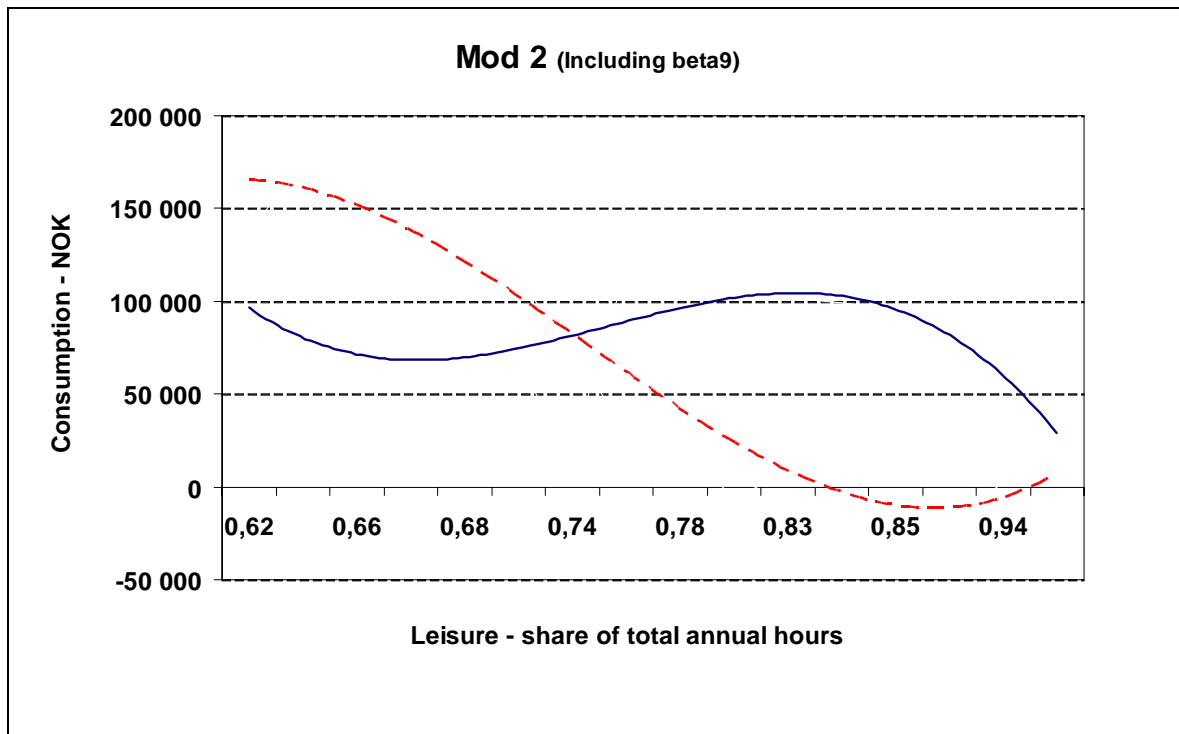
Fig. 4. Indifference curves for model 1.

Mod 1 (Excluding beta9_a=0)

Consumption - NOK vs *Leisure - share of total annual hours*

According to model 1 the needed compensation in terms of consumption for working one more hours is almost the same irrespective of hours. Furthermore, when the leisure share of total hours becomes very high, consumption valued in NOK approaches zero, and even becomes negative for women. This seems highly unreasonable.

From the predictions of model2 it seems that men and women have an opposite development in preferences. At low levels of leisure, implying high levels of labour hours supplied, women demand less and less consumption compensation for one hour more of work as leisure becomes even lower. This is highly unlikely in my view, as women in general most certainly treasure consumption as much or more than men. Again, the unreasonable feature is suggested namely that required consumption compensation for one hour more of work for women becomes negative at high levels of leisure. For men the indifference curve seems reasonable at low levels of leisure, but has an unlikely development at levels of leisure above 70% of total annual hours.

Fig. 5. Indifference curves for model 2.

Model 3 suggests concave indifference curves, i.e. the complete opposite of standard expectations of individual preference structures. According to model 3, individuals demand increasingly more consumption compensation for one more hour of work as the level of leisure becomes higher. At low levels of leisure, the compensation required even is negative.

Fig 6. Indifference curves for model 3.
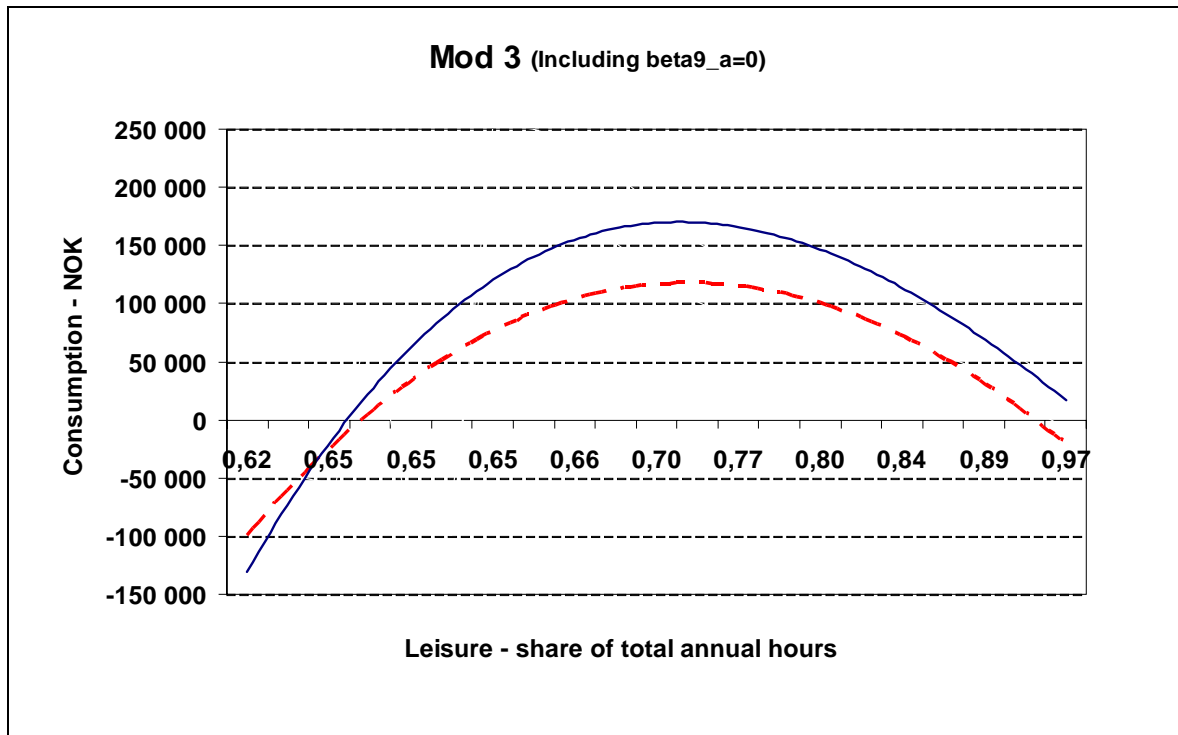
Considering the indifference curves estimated by the flexible models I think it is fair to say that they fail to predict in a satisfying way individual behaviour when it comes to consumption and labour supply decisions in an environment with a black labour market.

5.4.2. Predictions

Predicted behaviour is compared to the observed behaviour, as reported in the survey. As expected the structural model seems to have an obvious advantage, producing numbers much closer to expected values than the flexible ones. Model 4 seems to best estimate overall actual behaviour. Of the two models predicting the choice of white hours, the model based on only the honest respondents, model 5a, most closely resembles reported behaviour. As demonstrated by

the predicted values, the flexible models seem to have only a weak, if any, ability to predict actual behaviour in this setting.

Table 10. Observed values and model predictions.

Variables	Observed	Predicted					
		Mod 1 Excl beta9_a=0	Mod 2 Incl beta9	Mod 3 Incl beta9_a=0	Mod 4	Mod 5a White-honest	Mod 5b White-all
P(H)	0,888	0,466	0,572	0,490	0,888	1,000	1,000
P(E)	0,112	0,534	0,428	0,510	0,112		
(L$_H$|H)	1733	1961	1948	1964	1880	1878	1890
(L$_H$|E)	1768	1382	1470	1456	1720		
(L$_E$|E)	79	278	194	242	300		
L$_H$	1736	1724	1793	1750	1864		
L$_E$	9	148	84	126	34		

5.4.3. Goodness of fit: log likelihood values and McFadden's rho (ρ^2)

The McFadden's rho-squared-values for model 4 are as earlier mentioned higher than the ones for models 1 to 3. This supports the picture my analysis so far has suggested, namely that the structural model 4 is better suited than the suggested flexible models in the task of predicting individual utility maximizing behaviour when it comes to preferences regarding consumption, labour supply and tax evasion. It is worth noting, however, that the relatively high goodness of fit values that the flexible models produce become irrelevant when one considers the fact that these models predict unrealistic results for individual utility maximizing behaviour.

Table 11. McFadden's Rho-squared for all the models.

McFadden's Rho:

Mod4	Mod5a	Mod5b
0,552	0,441	0,369

Mod1	Mod2	Mod3
0,366	0,427	0,395

5.5.   Conclusion

With my analysis I have wanted to find out if the type of Box-Cox structural utility model presented here is a better representation of individual utility maximizing behaviour than a flexible model, represented by a polynomial of degree 3. Considering the models estimates, predictions and goodness of fit values it seems that the structural model I have worked with has a clear advantage over the suggested flexible models for predicting individual behaviour when it comes to consumption and labour supply decisions in an environment with a black labour market.

6. References

Allingham, M.G. and A. Sandmo (1972): "Income Tax Evasion: A Theoretical Analysis", *Journal of Public Economics*, 1, 323-338.

Ben-Akiva, M (1973): "Structure of Passenger Travel Demand Models", Ph.D. Dissertation, Department of Civil Engineering, MIT, Cambridge, Mass.

Ben-Akiva, M. and S. Lerman(1979): "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility". In D. Henscher and P. Stopher (eds): Behavioral Travel Modelling. Croom Helm, London.

Dagsvik, J. K. and Strøm, S. (2005): "Sectoral labor supply, choice restrictions and functional form", forthcoming in *Journal of Applied Econometrics*

Jørgensen, O, Ognedal, T and Strøm, S. (2005): "Labor supply when tax evasion is an option", Memorandum, Department of Economics, No 6.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*, 47, 263-291.

Markeds- og Mediainstituttet A/S (2003): *Det skjulte arbeidsmarked (postal), 2003.*

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior", Zarembka, P. (ed.) Frontiers in econometrics, New York and London, *Academic Press*, 105-142

McFadden, D. (1978): Modeling the choice of residential location, in A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, eds, *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75-96.

Quiggin, J. (1982): "A theory of anticipated utility", *Journal of Economic Behavior and Organization* 3(4), 32343

Quiggin, J. (1993): "Generalized expected utility theory: the rank-dependent model", Boston: Kluwer, XII, 208 s

Rice, J. A. (1995): *Mathematical Statistics and Data Analysis*. Second edition. Duxbury Press

Romer, D 2001: *Advanced macroeconomics*. Secon edition. McGraw-Hill/Irwin, New York

http://water.usgs.gov/pubs/of/2003/ofr03500/pdf/ofr03-500.pdf

http://www.bwk.tue.nl/urb/p/Danielle/acsp99.html

Statistics Norway: Consumer expenditure surveys: http://www.ssb.no/fbu/

7. *Appendix*

Questions asked in the questionnaire
The respondents were asked to cross out answer-alternatives that vary across the questions. These alternatives are not shown here, but are available upon request.

Q.1. Gender

Q.2. Age

Q.3. Number of children living in the house

Q.4. Marital status

Q.5. Does your spouse have income generating work, and if so, how many hours?

Q.6. Education in years

Q.7. Occupational status (wage worker, self-employed, unemployed, retired, etc)

Q.8. Hours of work last week in the regular economy

Q.9. Hourly wage rate in main occupation

Q.10. Annual, net income (after tax) in main occupation

Q.11. Annual gross income in main occupation

Q.12 Occupation by industry

Q.13. Do you receive other income than wage income such as social security benefits/unemployment benefits/capital income?

Q.14.What is your tax rate for overtime work, the marginal tax rate in percent?

Q.15 How much tax do you pay in percent of your total annual gross income?

Q.16.What do you think is the attitude among people with respect to receive payment for work that is not reported to the tax authorities? Do you think it is accepted/accepted to some extent/not accepted/don't know

Q.17. Have you ever been engaged in non-reported income activities?

Q.18. If so, what kind of activities was it?

Q.19. If you had the opportunity to receive income without having to report it to the tax authorities, would you then accepted such income?

Q.20. If you don't report income to the tax authorities, how large do you think the chance (percent) is that you would be caught?

Q.21. If you do not report income to the tax authorities, say NOK 20 000, and you are caught; you have to pay a penalty tax in addition to the regular tax on the nonreported income. How large do you think this penalty tax rate is (percent)?

Q.22. During the last 12 months, have you received compensation for work that has not been reported or will not be reported to the tax authorities?

Q.23. Approximately how many hours of non-reported work have you done during the last 12 months?

Q.24. At the last tax declaration; what was the total annual income from work and capital income that you did not report?