CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

1999

Estimating the Welfare Cost of Taxation in a Labour Market with Unemployment and Non-Participation

Vincent Hogan
University College Dublin

Working Paper

WP99/2

February 1999

DEPARTMENT OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Estimating the Welfare Cost of Taxation in a Labour Market with Unemployment and Non-Participation[*]

Vincent Hogan
Department of Economics
University College Dublin

January 29, 1999

Abstract

The standard public finance analysis of the welfare cost of labour income taxation is based on the estimation of labour supply functions that treat unemployed individuals as non-participants. This paper applies econometric models of multinomial discrete choice to the labour market, explicitly allowing individuals to be in any of three possible states (employment, unemployment and non-participation). Based on these estimates, we present calculations of the dead-weight loss of taxes, which turn out to be much larger than those suggested by the standard literature.

[*] I would also like to thank the following for their advice and support: Colm Harmon, Jerry Hausman, Jim Poterba and Olivier Blanchard Naturally, the errors (all of them) are mine.
Email abuse to vincent.hogan@ucd.ie

1 Introduction

The calculation of the welfare cost of labour market taxes has been a focus of research in public economics for at least two decades. This level of interest is not surprising. Modern governments raise such a high proportion of their revenue from income and social security taxes that the welfare costs of these taxes are of crucial interest to policy makers. What is surprising, however, is that most of the empirical analysis has tended to avoid the issue of unemployment. The focus has been on the estimation of hours equation over samples consisting of the employed workers only. Some authors, especially when analyzing women's labour supply, did allow for non-participation.[1] In general, however, little attempt has been made to account for the unemployed. This paper calculates the welfare cost of labour market taxes in a frame-work that explicitly allows for unemployment and non-participation.

Within the frame work of the traditional literature there are two possible ways of dealing with those individuals observed to be unemployed. One can either include them in the estimation sample and treat them as having supplied zero hours or, alternatively, one can exclude them from the sample, using only the employed to estimate labour supply elasticities. Ham (1982) showed that both these approaches resulted in biased estimates of the labour supply function. If the unemployed are included in the estimation sub-sample, and if they are truly constrained, then the difference between observed and desired hours will be incorporated into the residual. Thus the residual will be correlated with the independent variables rendering least squares estimates of the labour supply function biased and inconsistent. Alternatively, if the unemployed are dropped, sample selection bias is introduced. Unless one is prepared to believe that unemployment is a random event whose occurrence is uncorrelated with any variables that may influence tastes for work, estimation over a sample consisting of only the employed, will lead to biased results.

Ham (1982) dealt with this problem by excluding the unemployed from the sample and correcting for the resulting selection bias using Heckman (1979) procedure. However, he did not offer a consistent treatment of non-participants, those who reported zero hours but who were not unemployed. Non-participants may have tastes for work that are systematically different from either those of the employed or the unemployed. Thus treating them as being like the employed (with zero hours "supplied") or like the unemployed, would lead to biased estimates. Similarly, excluding them from the sample would lead to another selection bias.

In this paper, I apply multinomial discrete choice econometric models to labour market data from the UK. These models explicitly allow an individual to occupy any of three labour market states (employment, unemployment, non-participation). Issues of sample selection do not arise because these models can be estimated over the entire sample.

Two classes of models are used. The first class assumes that the individual

[1] See Mroz (1987) for an example.

can choose to be unemployed. In this sense these are models of voluntary unemployment. Unemployment is still distinct from non-participation, however. Suppose, for example, that an individual turns down an offer of a job at a low wage and opts instead to remain searching. His unemployment is voluntary, in the sense that he could have accepted the low wage job or have dropped out of the labour market. This setup leads to Multinomial Logit (MNL) and Nested Multinomial Logit (NMNL) estimators.

The second class of models treats unemployment as being completely involuntary. Individuals can choose whether or not to participate. Conditional on participation, they get a job with some probability. When making the participation decision individuals are aware that participation does not guarantee employment. This setup leads to a what may be described as a "nested probit" estimator.

These estimates confirm some reasonable hypotheses regarding the labour market. Most particularly, it seems that employment, unemployment and non-participation are all qualitatively different states. Formal tests of the hypothesis that the unemployed have the same preferences as either the employed or the nonparticipants are overwhelmingly rejected. Thus the results confirm those of Flinn & Heckman (1983). Any analysis of labour supply that fails to take this into account could be misleading.

The Deadweight Loss (DWL) of taxes can be calculated from these estimates using the methods of Small & Rosen (1981) and McFadden (1986). The resulting values of the DWL, in both the voluntary and involuntary unemployment models are considerably higher than those found in much of the literature.

A possible explanation for these results is that the estimates presented here capture the effect of wages (and taxes) on participation. The estimates in the traditional literature report the effects of wages on hours supplied, conditional on participation. If there are fixed costs of entry to the labour market, then we might expect the participation elasticity to be higher than the hours elasticity.

Although not the original purpose of this research, it is interesting to note that these results may help shed some light on the controversy regarding the alleged bias of the Non-Linear Budget Set (NLBS) model of Hausman (1985). The results presented here are similar in magnitude to those calculated using theNLBS method by Hausman (1985). This is despite the fact that this paper I treat the budget constraint as being linear. Some authors[2] have suggested that the NLBS results are due to the estimation method imposing high compensated elasticities on the data. It may be, however, that the difference in the results is due to the different treatment of the non-participants. The linearized method is typically estimated over the employed, so it returns the (low) compensated elasticity of hours conditional on participation. The NLBS method, being a generalized version of the Tobit procedure, easily accommodates the non-participants (but not the unemployed). Therefore it would return an average of the participation and hours elasticity. The difference in the results of the two procedures could be entirely due to a high elasticity of participation.

[2] See MaCurdy (1992) for example.

Clearly further research is required before a definitive statement can be made on this issue, but the results presented here are suggestive.

The models estimated in this paper represent an improvement over the traditional literature in so far as they allow for both non-participation and unemployment. They are still lacking in one significant aspect, however. In common with much of the literature, they are partial equilibrium. I do not model the aggregate constraint that generates unemployment. I assume that there is some inefficiency operating behind the scenes that prevents the labour market from clearing. I do not investigate how this inefficiency interacts with the tax system. This interaction could have important implications for social welfare. The imposition of even a small tax on an already distorted economy could have first order welfare effects increasing even more the welfare cost of taxation. Suppose, for example, that unemployment is the result of inefficient matching of job seekers and vacancies. The imposition of a tax on labour income could reduce an individual's search effort. Apart from the direct effect on the individuals welfare, this would also reduce other individuals' and firms' welfare via the search externality: one persons reduced search effort reduces the effectiveness of all other individuals' job search and firms' efforts to fill vacancies. Thus, the search externality magnifies the welfare cost of taxation.

Alternatively, the two sources of inefficiency may actually counteract each other. Continuing the search example, suppose that workers face a cost of participation in the labour market and/or a search cost. In this case, anything that forces workers to leave the labour market, will save them the search cost. Thus the cost of taxation would be less than estimated here.

As yet another alternative, one could view unemployment as a job queue in a ranking model i.e. where firms have multiple applications for their vacancies. In this economy, anything that causes an individual to leave the queue, raises the welfare of those remaining on it, without hurting the firms who receive multiple applications. In this case the search externality reduces the welfare cost of taxation.

The point being made here is that it is possible to envisage a multiplicity of interacting effects. The dead-weight loss calculations presented here do not account for any of these effects (if they exist). It is not possible to sign the bias that results from this omission, as the examples given above indicate, it could go either way. The standard models of competitive general equilibrium used in public finance to examine the issue of tax incidence are obviously unable to resolve this issue. Its resolution requires further research into the incidence of tax in non-competitive models. Correcting this omission is an obvious topic for future research.

The paper is arranged as follows. Section 2 discusses the previous approaches to calculating the welfare cost of labour income taxation. Section 3 presents the econometric models used in this paper. Section 4 discusses the data. Section 5 presents estimates of the two classes of models and calculates the deadweight loss of taxes. Section 6 concludes.

2 Previous Approaches to labour Supply

This section briefly reviews the previous approaches to the estimation of models of labour supply. Consider estimating a typical labour supply function such as (1) over a sample consisting of the employed workers only, using Two Stage Least Squares (2SLS).[3] The estimates can be used to calculate the deadweight loss (DWL) of taxes using the Harberger triangle formula.[4]

$$H_{it} = \alpha + \beta W_{it} + \gamma Y_{it} + \delta X_{it} + \varepsilon_{it} \tag{1}$$

The results (shown in table 1) are consistent with theory in so far as the estimated Hicksian wage elasticity is positive. The size of the elasticity is within the range reported by other authors for the UK, although it is at the higher end of the range.[5] Furthermore, the estimated DWL of taxes as a percentage of the revenue raised, is within the range reported by many other researchers. But it much smaller than those reported by Hausman (1985) using the method of Non-Linear Budget Sets (NLBS).

Ham (1982) noted, however, that least squares applied to (2) will almost certainly be biased, if we admit the possibility that there is unemployment. In order to illustrate the point, suppose that equation (2) is the true labour supply function i.e. H^*_{it} is the hours of work desired by the individual, where X is a vector of independent variables and ε is a normally distributed error.

$$H^*_{it} = X_{it}\beta + \varepsilon_{it} \tag{2}$$

Suppose that we try to estimate (2) using a sample consisting of both the employed and the unemployed. In effect, we would be estimating equation (3) where $u(Z_{it})$ is the probability that an individual i is unemployed at time t and H_{it} is the actual observed hours worked, with $H_{it} = 0$ for the unemployed. Equation (2) incorporates into the residual any difference between observed and desired hours caused by unemployment.

$$H_{it} = X_{it}\beta + \eta_{it} \tag{3}$$

$$\eta_{it} = \begin{Bmatrix} \varepsilon_{it} & w.p & 1 - u(Z_{it}) \\ \varepsilon_{it} - H^*_{it} + H_{it} & w.p & u(Z_{it}) \end{Bmatrix}$$

If the unemployed are not truly constrained, so that $E(H^*_{it} - H_{it}) = 0$, then there is no problem (other than heteroscedasticity). They could be treated as censored observations and equation (2) estimated using the Tobit procedure. If they are constrained, however, it is clear that $E(\eta|X) = -H^* u(Z) = -X\beta u(Z)$, then the residual is correlated with the independent variables and least squares on (3) will be biased. The are two sources of bias. Firstly, even if unemployment

[3] The instruments used were education and household compositin variables together with regional dummies and the regional unemployment rate. See the discussion in section 4.

[4] $\frac{DWL}{R} = \frac{1}{2} t \varepsilon^h$ where ε^h is the Hicksian elasticity of labor supply and t is the ad-valorum tax rate.

[5] See Blundell (1993) and Pencaval (1986).

is random i.e. $u(Z)$ is a constant, any change in X will change the size of the constraint faced by the unemployed. This biases the least squares estimate of β towards zero. Secondly, if some element of Z is also in X, then the extent to which the constraint is binding in also changes. Least squares estimation of equation (3) incorporates both these effects into the estimate of β.

An example will help clarify the issue. Consider a variable measuring an individual's educational achievement. It is plausible that this variable affects an individuals taste for work. It is equally plausible that education also affects an individual's chances of becoming employed. In this case, if we estimate (2) over all individuals, employed and unemployed, the residual will be correlated with one of independent variables generating inconsistent estimates of education on individuals' desired hours.

Ham (1982) notes that the alternative procedure, excluding the unemployed from the estimation sample, is also likely to be biased. This is a standard sample selection problem. Least squares estimation of (2), over the sub-sample of the employed, will be consistent only if $E(\varepsilon|\text{employed}) \neq f(X)$. If unemployment is a random event, so that the unemployed are not systematically different from the employed, then this condition is satisfied. We expect, however, that the probability of being unemployed is a function of the same sort of variables that enter a labour supply equation (education, for example). Intuitively, if the unemployed tend to be those with less education, their exclusion from the sample will bias estimates of the influence of education on tastes for work. Ham (1982) corrected for this bias using the selection model of Heckman (1979) to estimate a labour supply equation over the sample consisting of employed individuals only.

The Heckman procedure can deal successfully with the problem of unemployment, but does not account for the possible distinction between the unemployed and non-participants. In fact we face a similar set of issues regarding non-participants as we did regarding the unemployed. As in the case of the unemployed, we can either include or exclude non-participants from the sample used to estimate the model. In either case the estimates are likely to be biased.

If we believed that non-participants were randomly selected, then we could safely exclude them from the estimation sample. If we believed that they were not significantly different from the employed, then we could treat them as censored observations and estimate a Tobit type model. If we believed that they were not significantly different from the unemployed we could simply re-estimate the Ham (1982) model treating them as such. If we believe, however, that the non-participants are, potentially, systematically different from both the employed and the unemployed, then we ought to account for them separately. This is most clearly done by estimating multinomial models of discrete choice. In this paper, I apply multinomial discrete choice econometric models to labour market data from the UK. These models explicitly allow an individual to occupy any of the three labour market states (employment, unemployment, non-participation). Issues of sample selection do not arise because these models can be estimated over the entire sample.

3 An Econometric Model of Unemployment

It is clear from section 2 that we need an econometric procedure that is flexible enough to allow for the three labour market states. Of crucial importance to this how unemployment is modeled. I adopt two polar assumptions. In the following sub-section I propose estimators based on the assumption that unemployment is voluntary (in a sense to be made precise shortly). In the subsequent section I propose an estimator based on the opposite assumption i.e. unemployment is completely involuntary. In either case the Random Utility Model (RUM)[6] is a convenient framework within which to consider the alternative procedures.

A RUM model of labour market choice is given by equation (4). An individual, i, assigns (indirect) utility (U_{it}^j) to each of the three labour market states in each time period. Utility consists of a deterministic component (V_{it}^j) and an stochastic component (ε_{it}^j) where $j, k \in \{e, u, n\}$ indicate employment status (employed, unemployed and non-participant). The deterministic component is a "linear in parameters" function of X_{it}^j, a matrix of independent variables at time t, and α^j a vector of individual specific effects. The stochastic components have a joint distribution $F(\varepsilon_{it}^e, \varepsilon_{it}^u, \varepsilon_{it}^n)$ with a covariance matrix Σ. For convenience, I will assume that the covariance matrix is the same for all individuals (homoscedastic) and for all time periods. The stochastic component of utility can be interpreted as either true randomness in preferences across time and individuals, or as being that component of utility that is not observed by the econometrician. Specifying the functional form of F generates a particular econometric model.

$$\begin{aligned} U_{it}^j &= V_{it}^j + \varepsilon_{it}^j \\ V_{it}^j &\equiv \alpha_i^j + X_{it}^j \beta^j \end{aligned} \tag{4}$$

Equation (4) implies that each of the three labour market states may induce different levels of utility. In particular, being unemployed is not necessarily the same as being a non-participant. In this manner the model already departs from much of the traditional analysis.

3.1 Voluntary Unemployment

In the standard RUM the individual chooses whichever state provides the highest utility. So, for example, the probability that an individual chooses employment at time t, conditional on individual effects (α_i), is given by equation (5). The other probabilities are defined symmetrically.[7]

$$\begin{aligned} \Pr(E_{it}|\alpha_i) &= \Pr\{U_{it}^e > U_{it}^j \text{ where } j = (u, n)\} \\ &= \Pr\{\varepsilon_{it}^j < V_{it}^e - V_{it}^j + \varepsilon_{it}^e \text{ where } j = (u, n)\} \end{aligned} \tag{5}$$

[6] See McFadden (1986) for a detailed discussion of the RUM.

[7] We cannot identify all the parameters of the model, because we observe only the demand functions and not the utility functions directly. If a variable is constant across different choices for the same individual then a normalization is necessary. In what follows I normalize U^n to be zero. Also I assume for simplicity that $\alpha^e = \alpha^u$. Attempts to estimate α^e and α^u separately failed.

Equation (5) is anything but innocuous. It implies that individuals may choose to be unemployed. In fact they will do so if $U^u_{it} > U^e_{it}$ and $U^u_{it} > U^n_{it}$. In this sense unemployment is voluntary. This does not mean, however, that unemployment is the same as non-participation. The two states could differ in terms of entitlement to public assistance or in the degree of effort exerted in searching for employment. Unemployment is voluntary however, in the sense that an unemployed individual could have accepted the low wage job or have dropped out of the labour force. An econometric procedure based on (5) is strictly correct only to the extent that we can believe that unemployment is truly voluntary. This may strike some as being unreasonable, which is why, in the next sub-section, unemployment is assumed to result entirely from the imposition of a constraint.

With that caveat in mind we can estimate a model based on (5) by specifying a functional form of F. If we denote the partial derivative of F with respect to ϵ^e by F_e, then (5) can be re-written as

$$\Pr(E_{it}|\alpha_i) = \int_{\epsilon^e=-\infty}^{\infty} F_e(\epsilon^e_{it}, \epsilon^e_{it} + V^e_{it} - V^u_{it}, \epsilon^e_{it} + V^e_{it} - V^n_{it})\, d\epsilon^e_{it} \tag{6}$$

The integrand in (6) is the probability that ϵ^e equals a particular value and that the other disturbances are such that E will be chosen. By integrating over the support of ϵ^e we get the probability that E is chosen (conditional on α_i). The probabilities of the other employment states are derived symmetrically.

In order to control for the individual (random) effects, we integrate them out of equation (6). We then calculate equation (7), the log likelihood of the sample of N individuals over T periods, where d^j_{it} is a dummy variable equal to one if individual i is observed in state j at time t, and ϕ is the p.d.f. of the random effects. In general the first, and second, order conditions of (7) do not have convenient analytical forms. Maximization of the likelihood will have to be done numerically.

$$\ln L = \ln \prod_{i=1}^{N} \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{T} \sum_{j=e,u,n} \left[d^j_{it} \Pr(J_{it}|\alpha_i) \right] \right\} \phi(\alpha)\, d\alpha \tag{7}$$

Before proceeding, it is worth making clear why fixed effects estimators are not used. In general, these are not available for non-linear models. Because of the non-linearities, the fixed effect cannot be differenced out as with the least squares "within-groups" estimator. Therefore, it must be estimated as an incidental parameter. This yields consistent estimates only if there are sufficient observations through time foreach individual. This is not true in most datasets. Furthermore, because the MLE of the coefficients on the independent variables and the incidental parameter are jointly determined, any inconsistency will be passed on to the estimates of the other coefficients.

Small & Rosen (1981) and McFadden (1986) provide method of calculating the dead-weight loss in the for discrete choice models. They show that in the case of a RUM such as (4) that the probability function (6) can be treated as being the demand function and that many of the standard results of consumer theory for continuous goods hold also for discrete choice. In particular, the Compensating Variation (CV) of any price change can be calculated by integrating the area under the compensated demand curve, to get equation (8), where μ is the marginal utility of labour income ($\mu = \delta V^e / \delta W$).[8]

$$CV = \frac{1}{\mu} \int_{V_0^e}^{V_1^e} P^e(V^e)\,dV^e \tag{8}$$

The Dead Weight Loss (DWL) of a tax is given by the CV less the tax revenue measured at the compensated point. Note that the scale of Compensating Variation is important. It is straight forward to show that its scale will be the same as that of the other income variable, which here is pounds sterling per month. Thus the CV calculation will tell us the amount an individual would have to be paid on a monthly basis in order that he be no worse off as a result of the imposition of the tax.

3.1.1 Multinomial Logit (MNL)

Following McFadden (1974), assume that the ϵ in (4) are i.i.d. extreme value[9], then the probability of observing an individual in state $j \in \{e; u; n\}$ is given by (9).

$$P(J_{it}|\alpha_i) = \frac{\exp(V_{it}^j)}{\exp(V_{it}^e) + \exp(V_{it}^u) + \exp(V_{it}^n)} \tag{9}$$

By maximizing the log-likelihood in (7) with $P(J_{it}|\alpha_i)$ given by (9) we can estimate the MNL model controlling for individual effects.

It is worth noting at this point that Chamberlin (1980) proposed a fixed effects estimator for the MNL model. Unfortunately this procedure effectively drops from the sample all those whose status has not changed over the period of the sample. This would tend to bias the estimates if, for example, those who were employed for the entire duration are systematically different from those who were not. Therefore, when I estimate a MNL model I adopt the alternative to the Chamberlin method, treating the individual effects as being random, drawn from a normal distribution, and integrate them out, as in equation (7). This allows the MNL to be estimated over the entire sample without the loss of any class of observations

The expression in (8) for the compensating variation of a price change has a convenient closed form in the case of the MNL model, reducing to equation

[8] In order to derive (8) we must assume that the marginal utility of income is independent of wages. This is true in the models estimated here because utility is linear in the income term. In this case the Marshallian concept of consumer surplus coincides with the compensating variation.

[9] $F(\epsilon^e; \epsilon^u; \epsilon^n) = \exp(-[\exp(-\epsilon^e) + \exp(-\epsilon^u) + \exp(-\epsilon^n)])$

(10) which is also derived in McFadden (1986).[10]

$$CV = \frac{1}{\mu}\left[\ln\left\{\sum_{j\in\{e,u,n\}}\exp(V^j)\right\}\right]_{V_0^j}^{V_1^j} \tag{10}$$

3.1.2 Nested Multinomial Logit (NMNL)

The assumption inherent in the MNL model, that the errors are i.i.d. is not reasonable in many situations. It imposes restrictions on the cross price elasticities that are often unrealistic.[11] The model effectively assumes that all the alternatives are equally similar, once we have controlled for observables. In the context of the labour market, we might expect, for example, that employment and unemployment are similar, because both imply attachment to the labour force. Therefore, their random components will be correlated. Alternatively, we may think, that it is unemployment and non-participation that are similar as both are types of "non-work". Therefore it is the random components of utility of non-participation and utility of unemployment that should be correlated.

The NMNL model, due to McFadden (1986) was designed precisely to deal with this problem.[12] The probability of observing an employed individual is given by (11), with the other probabilities being defined symmetrically.[13]

$$P(E_{it}|\alpha_i) = \frac{\exp(V_{it}^e)\{\exp(V_{it}^e) + \exp(V_{it}^u)\}^{\rho-1}}{\{\exp(V_{it}^e) + \exp(V_{it}^u)\}^{\rho} + \exp(V_{it}^n)} \tag{11}$$

The parameter ρ represents the degree of dissimilarity between the unobserved components of the utility of employment and unemployment.[14] When ρ is one, all the unobservables are equally similar, equation (11) reduces to equation (9) and the NMNL reduces to the MNL model.

The model is estimated by substituting (11) into the likelihood function (7). The model is more easily understood, however, if it is expressed in terms of tree structure of sequential conditional probabilities. We can think of estimating

[10] Hausman et. al. (1997) used (10) to calculate a portion of the welfare cost of Exon Valdez oil spill.

[11] In the literature this restriction is often refered to as "Independence of Irrelevant Alternatives" (IIA).

[12] The Multinomial Probit Model would also allow us to avoid the IIA problem. Keane (1992) showed, however, that while the covariance parameters are formally identified, they can easily be mimicked by the coefficients on the independent variables. Thus, in practice estimation may fail, unless we can impose restrictions which may help separate movement in the independent variables from movements in the covariance parameters. I failed to estimate the model successfully with any economically sensible restrictions other than IIA! As one would expect, these results were similar to those of the MNL model and are not reported here.

[13] For the NMNL model the joint distribution of the errors is given by $F(\varepsilon^e;\varepsilon^u;\varepsilon^n) = \exp(-[\exp(-\varepsilon^e) + \exp(-\varepsilon^u)]^{\rho} + [\exp(-\varepsilon^n)]^{\rho})$

[14] Ben-Akiva and Lerman (1985) show that $\rho = \sqrt{1 - corr(\varepsilon^e;\varepsilon^u)}$

the model in two stages.[15] Suppose that labour market decisions are structured as follows. Individuals choose between participating in, or dropping out of, the labour force. This is the standard labour supply, or participation, decision. Then, conditional on participation, individuals choose either to become employed or unemployed.

$$P(L_{it}|\alpha_i) = \frac{\exp(\mu\lambda_{it})}{\exp(V_{it}^n) + \exp(\mu\lambda_{it})} \tag{12}$$

$$P(E_{it}|L_{it};\alpha_i) = \frac{\exp(V_{it}^e)}{\exp(V_{it}^e) + \exp(V_{it}^u)} \tag{13}$$

$$\lambda_{it} \equiv \ln(\exp(V_{it}^e) + \exp(V_{it}^u))$$

Each decision can be modelled as a standard binomial logit. The first stage "participation" logit (12) has an extra regressor λ_{it}, known as the "inclusive value". This regressor can be thought of as a composite utility index which summarizes the outcome of the second stage "employment" logit (13) i.e. it represents the utility of participation. The coefficient on the inclusive value (μ) is a measure of the similarity of U and E because it applies equally to both V_{it}^e and V_{it}^u in the regression.

McFadden (1986) notes that a sufficient condition for the estimated model to be consistent with utility maximization is that $\mu \in [0;1]$.[16] If the estimated model was found to be inconsistent with utility maximization then welfare calculations based on these estimates would be meaningless. According to McFadden (1986), if the estimate of μ is outside the permissible range the fault may lie with the specification of the decision tree i.e. in our example it may be unemployment and non-participation that are similar and not employment and unemployment. We estimate the model using both structures.

Once we have estimated the NMNL model satisfactorily, we can perform welfare calculations as in the MNL model. For the case of the NMNL the expression for the Compensating Variation of a price change given by (8) has a closed form, reducing to equation (14).

$$CV = \frac{1}{\theta}\,\frac{1}{\mu}\ln\{\exp(V^n) + \exp(\mu\lambda)\}\Big|_{V_0^j}^{V_1^j} \tag{14}$$

3.2 Involuntary Unemployment

The estimators of the previous section were based on the idea that unemployment was voluntarily chosen but still different from non-participation. In this

[15] The sequential estimator can be used to produce consisten, but inefficient, estimates of the model (see McFadden, 1984).

[16] To see this, one can differentiate (11) with respect to V^u. For the model to be consistent with utility maximization, this derivative must be negative i.e. the probability that the individual chooses employment must fall when the utility of an alternative increases. The derivative will be negative for all values of V^j if $\mu < 1$. It may be negative for certain values of V^j even if $\mu > 1$. Therefore $\mu < 1$ is a sufficient, but not necessary, condition for utilty maximization.

section we take the opposite position and assume that unemployment is completely involuntary. Unemployment is the result of a constraint. The individual can do nothing to alter the probability of being employed once he decides to participate. They only choice that exists is between participation and non-participation. Suppose, that the labour market decisions are structured as follows: Individuals choose between participating in, or dropping out of, the labour force. This is the standard supply decision. Then, conditional on participation, individuals either get a job or become unemployed. The second stage is the result of the imposition of an outside constraint. This is the sense in which unemployment is involuntary. Individuals make their participation decision knowing that participation does not guarantee employment. The probability that an individual gets a job if he participates is modeled as a function of exogenous variables that are outside the control of the individual.[17] For example, a youth with a low level of education, living in a high unemployment area has a low probability of being employed. Knowing this, he may be less willing to seek work at any given wage level.

Formally this structure can be modeled as follows: Individuals choose between participation, which yields indirect utility U^l_{it}, and non-participation which yields U^n_{it}: The utility of non-participation is as defined previously. The utility of participation is modeled as a function of the payoff if the individual secures employment, the payoff if he becomes unemployed and the probability that participation will yield employment. A convenient and logical specification is to model the utility of participation as the expected utility of the gamble that participation will yield employment with probability p_{it}. This is shown in equation (15) where the random variables, U^e_{it}, U^u_{it} and U^n_{it} are defined in (4) and $\eta_{it} \equiv p_{it} \cdot \varepsilon^e_{it}(1 - p_{it}) \cdot \varepsilon^u_{it} + \varepsilon^n_{it}$

$$\begin{aligned} U^l_{it} &= p_{it} \cdot U^e_{it} + (1 - p_{it}) \cdot U^u_{it} \\ &= p_{it}(V^e_{it} - V^u_{it}) + V^u_{it} + \eta_{it} \end{aligned} \tag{15}$$

The individual will choose participation if $U^l_{it} > U^n_{it}$. Conditional on participation he will get a job with probability p_{it}. Therefore the probability of observing the individual in each of the three states is given by (16). Substituting (16) into (7) will enable FIML estimation of the model controlling for random effects.

$$\begin{aligned} P(L_{it}|\alpha_i) &= P(U^l_{it} > U^n_{it}) \\ P(E_{it}|\alpha_i) &= p_{it} \cdot P(L_{it}|\alpha_i) \\ P(U_{it}|\alpha_i) &= (1 - p_{it}) \cdot P(L_{it}|\alpha_i) \\ P(N_{it}|\alpha_i) &= 1 - P(L_{it}|\alpha_i) \end{aligned} \tag{16}$$

The econometric model can be more clearly understood as a two stage model whose structure mirrors that of the labour market. First we estimate the probability of employment conditional on participation, p_{it} as a function of all the exogenous variables. This is a standard binomial discrete dependent variable

[17] Thus the analysis does not allow for any attempt by participants to improve their chances of finding employment by, for example, migrating to a different region or improving their educational qualifications.

model such as probit or logit. As no choice is involved, this part of the model has no RUM interpretation and no direct welfare implication. Then the fitted values from this regression can be used for p_{it} in estimating the binomial choice between participation and non-participation. This second stage has the standard RUM interpretation as discussed in the previous sub-section.

We can implement the expected utility model of equation (15) by specifying the errors to be i.i.d. standard normal random variables. The probability of participation is then given by $P(L_{it}) = P(U^l_{it} > U^n_{it})$.[18]

$$P(L_{it}|\alpha_i) = \Phi\left[\frac{p_{it}(V^e_{it} - V^u_{it}) + V^u_{it}}{\sqrt{Var(\eta_{it})}}\right] \tag{17}$$

As η is a sum of normal random variables of mean zero, it also is a normal random variable of mean zero. A potential problem arises with its variance, as this is a function of p_{it} which varies across individuals. Thus even if the ϵ are homoscedastic, the errors in the estimated model will be heteroscedastic. Heteroscedasticity in non-linear models leads to inconsistency as well as inefficiency (see Greene, 1993). Fortunately, the nature of the heteroscedasticity is clear in this case, thus we can model it directly. Assuming that ε are homoscedastic with identity covariance matrix, the variance of η is given by

$$Var(\eta_{it}) = 1 + 2 \cdot p^2_{it} - 2 \cdot p_{it}$$

Once we have estimated the model, we can then use the method of Small and Rosen (1981) to calculate the compensating variation of any taxation. The expression for the compensating variation of a price change in (8) must be modified to account for the fact that the individual now faces only a binomial choice between participation and non-participation. In this case the expression does not have a closed form and (18) must be evaluated numerically.

$$CV = \frac{1}{\mu}\int_{V^l_0}^{V^l_1} P^l(V^l)dV^l = \int_{V^e_0}^{V^e_1} \Phi\left[\frac{p_{it}(V^e_{it} - V^u_{it}) + V^u_{it}}{\sqrt{Var(\eta_{it})}}\right] dV^e \tag{18}$$

4 Data & Specification

The analysis is conducted using a British dataset, the British Household Panel Survey (BHPS),which covers four years from 1991-95.[19] This dataset has three particular advantages. Firstly, it contains information on the labour market status of respondents, enabling the difference between unemployment and non-participation to be clearly defined. Secondly, because the data is a panel, we will be able to control for individual specific (unobserved) effects. Thirdly, unemployment in Britain has been higher and more persistent than in the United States (though less so than in the rest of Europe). Therefore it is reasonable

[18] As before we normalize $U^n = 0$ and assume for simplicity that $\alpha^e = \alpha^u$.

[19] Table (8) reports the deinitions and, for the non-categorical variables, the summary statistics of the major variables used in the analysis. The variable rate_m is from the labor force survey.

to suspect that the treatment of unemployment may have more of an a®ect on measures of the dead weight loss of taxes in Britain rather than in the U.S.

I exclude women from the sample following the tradition in the labour supply literature which views the labour supply decision processes of the two sexes as being qualitatively di®erent. I also exclude from the analysis all those men who were employed but for whom no wage was reported and those over 70 years of age.[20] After dropping observations with missing values, we are left with a balanced panel of 1;866 men observed at each of the four annual waves.

The precise de¯nitions of employment status are obviously crucial for the question at hand. Unfortunately, the exact de¯nition of involuntary unemployment is problematic, as Ashenfelter (1978) pointed out. An economically sensible de¯nition, proposed by Ashenfelter (1978), would hold that an individual, A, is involuntarily unemployed if, and only if, two conditions are met. Firstly, there must exist another individual, B, identical to A in preferences and skills, who is employed. Secondly, A must be willing to take B's job under the same terms as B if it were to be o®ered. Obviously in practice it is impossible to implement this de¯nition because it is impossible to match people so closely.

Most statistically implemented de¯nitions of unemployment rely on individuals reporting their own employment status and then apply some consistency check to ensure that the constraint they claim to face is reasonable and that the unemployed individual is close to satisfying Ashenfelter's de¯nition. The standard OECD de¯nition requires that the unemployed individual has actively searched for work during the previous four weeks. If he has not done so, then he is judged to be a non-participant. In practice there is quite a lot of variation across countries in the precise de¯nition of the \standardized" unemployment rate.[21]

The OECD de¯nition seems too restrictive. Search e®ort is likely to be determined by the perceived probability of success as much as it is by the tastes of workers. The OECD de¯nition would tend to systematically exclude discouraged workers. These individuals may be as willing to accept the o®er of a job as are the other unemployed, but they are disillusioned with the e±cacy of job search and therefore no longer actively search. The problem is that we are trying to approximate a continuous variable (degree of attachment to the labour force or search e®ort) by a discrete division. There will always be a degree of ambiguity in the choice of where exactly we should draw the dividing line between unemployment and non-participation.

Jones and Riddell (1999) examined this issue in detail using Canadian data. They found that there is an important degree of heterogeneity among the group conventionally classi¯ed as non-participants: those who indicate that they desire to work exhibit distinctly di®erent behaviour from the remainder of the group.

I de¯ne as unemployed any individual who does not have a job but reported that he would work if he were o®ered a job. The precise de¯nition of employment status is as follows:

[20] Even at age 70, some 16% percent of respondants described themselves as being in the labour force.

[21] See OECD (1986)

- A person is de¯ned as employed (E) if he was working the week before the interview. For these individuals I observe hours worked strictly greater than zero and the usual net and gross wage.

- A person is de¯ned as unemployed (U), if he did not have a job the week before the interview but reported that he would take one if o®ered. Both the employed and the unemployed constitute the labour force.

- A person is considered to be \not in the labour force" (N) if he does not belong to either of the other two categories.

The sample unemployment rate is about 14% over the four years.[22] This makes it signi¯cantly higher than the o±cial unemployment rate for the U.K. over the same period, which was 10% or so. This re°ects the di®erence in the de¯nition of unemployment. Some individuals, who are classi¯ed as non-participants under the o±cial scheme, are classi¯ed as unemployed here. This probably results from a re-classi¯cation of discouraged workers as being unemployed rather than as non-participants.

In order to estimate the models of section three we need estimates of the market wage faced by the unemployed and the non-participants. Following the literature, I ¯rst regress the wage on the set of exogenous variables and use the ¯tted values as regressors in the models of section three.

The exogenous variables used here are similar to those used throughout the literature. They fall into three broad categories: human capital; household composition; and time and region ¯xed e®ects. The human capital variables are all related to education and occupation (educ, pasoc, tae). The household composition variables include marital status (marstat), the number of dependent children of various ages (k02y-k1618y) and a dummy for being head of household (hoh).

The wage equation was estimated by regressing (log) nwage on dummies created from the above exogenous variables.[23] We employ the Heckman (1979) procedure to deal with the sample selection problem. We ¯rst estimate a probit selection equation using all the exogenous variables together with interactions between time and region dummies. This probit is used to calculate the inverse mills ratio which appears as an additional variable in the wage equation. The results are shown in Table 6.

This procedure is not without controversy. Only the exogeneity of fathers occupation (pasoc), age and the time dummies are absolutely assured. One could argue that the human capital variables, the household composition variables and even the regional dummies are all the result of choices that are made jointly with the choice of labour market status and are therefore not exogenous.

The are two defences against this criticism. Firstly, it can be argued that, while these variables are formally the result of choices, those choices are su±-

[22] Unemployment rate is de¯ned as $\frac{U}{U+E}$

[23] An analagous gross wage equation was also estimated. It had the same speci¯cation as the net wage equation but with the log of gwage as the dependent variable.

ciently independent of the choice of employment status to enable the resulting variables to be treated as being exogenous in practice. For example, education may be thought to be endogenous because we choose a higher level of education in anticipation that this will enable us to earn higher wages, avoid unemployment etc. But it is also plausible to suggest that the choice of education is as much a result of available opportunities, parental encouragement, personal abilities etc. To the extent that this is so, the level of education can be considered to be exogenous. Similar arguments can be made in relation to the household composition variables.

The second defence is that the contribution of this paper to the literature, is not intended to be the improvement on the identi¯cation of previous models. Instead the focus of the paper is on the implications of unemployment for estimation of labour supply. It seems reasonable, therefore, to adopt standard practice in all other respects.

We also need to estimate the level of public assistance (bene¯ts) that would accrue were the current employed to change status. The U.K. bene¯ts system is quite complicated.[24] The unemployed and some non-participants can both be eligible for public assistance. Furthermore some low paid workers may also be eligible for the receipt of means tested bene¯ts. The system is too complicated to model directly.[25] Therefore we regress the log of bene¯ts on the exogenous variables correcting for selection using the heckit procedure. The idea is that the regressors, especially the household composition variables, will approximate su±ciently closely the eligibility rules for public assistance.

Finally, the issue of how to model the budget constraint must be decided. There is considerable controversy in the literature over whether it is appropriate to model the budget constraint directly as in Hausman (1985). Some authors suggest that the Non-Linear Budget Set (NLBS) procedure biases the results towards ¯nding large deadweight losses. Other authors maintain that it is the linearized method that it biased.[26] This decision is important because the results di®er substantially depending on which method is chosen, with the NLBS procedure usually generating much higher estimates of the cost of taxation.[27] For this paper it was decided not to use the NLBS methodology in order that nothing may detract from the focus of the paper on the implications of unemployment for measures of the welfare cost of taxation. We try to follow the traditional model in all its aspects, other than in its traetment of unemployment and non-participation. Any di®erence between the results presented here and the traditional literature can be attributed to the treatment of unemployment and not to the use of the NLBS method.

[24] For a survey see Atkinson and Micklewright (1991).

[25] Also modeling the bene¯t system directly would entail using the NLBS procedure. This is something we would rather not do for the reasons discussed in section 2.

[26] See Bloomquist (1995) for arguments in favour of the NLBS methodology and MaCurdy (1992) for arguments against.

[27] See Hausman (1985) for a comparisson.

5 Results

Table 2 shows the results of the estimation of both the MNL and NMNL models of voluntary unemployment. The first two columns show, for the MNL model, the coefficients of V^e and V^u, the indirect utility of employment and unemployment respectively. The coefficients of V^n, the indirect utility of non-participation, are normalized to zero. The model was estimated subject to the constraint that wages have no direct impact on unemployment and benefits have no direct impact on employment. In the context of the RUM, this restriction implies that the conditional utility of each state is dependent only on the "price" associated with that state. A Likelihood Ratio test of this restriction produces a χ^2 statistic of 0.68 which implies that we can not reject the restriction at any standard significance level.

The estimated coefficients take on plausible values. A higher wage tends to increase the relative probability that an individual is employed.[28] Similarly, higher benefits increase the probability that the individual will choose unemployment over employment or non-participation. The coefficient on the other income term indicates that higher non-labour income will tend to discourage participation in the labour market, either as a worker or as a job seeker. In general the coefficients on the polynomial in age were individually insignificant but collectively significant.[29] The presence of dependent children has no significant effect on the labour market participation of men. The coefficient on the dummy variable indicating that the individual is married, is positive as expected. Married men have greater attachment to the labour force than un-married men, and are more likely to be employed or to be seeking work and less likely to be non-participants.

These results suggest that unemployment and employment are very different states. Similarly we may wish to test is that the unemployed are the same as non-participants i.e. $\beta^u = \beta^n = 0$. If this hypothesis were true then we could explain the labour market behavior by estimating a traditional binomial logit model of participation. Again a likelihood ratio test of this hypothesis lead to its rejection confirming the result of Flinn and Heckman (1983).[30]

The estimates in Table 2 can be used to calculate the welfare cost of taxation as explained in section three. Table 3 shows this calculation and some of its intermediate steps, for the sample as a whole, and for quartiles of the distribution of the gross wage. The calculations use the estimates of the MNL model from Table 2. In order to understand how Table 3 is constructed, consider an average individual who earns $\$1,054$ per month and faces an average tax rate of 24%.[31]

[28] Some care should be taken when interpreting the coefficients in any multinomial model. They do not necessarily have the same sign as the marginal probabilities. In the case of MNL they represent the marginal effect of the covariates on the log of the odds ratio i.e. $\beta^e = \frac{\delta \ln(P^e/P^n)}{\delta X}$

[29] A Wald test of the hypothesis that all age variables should be dropped produced a χ^2 statistic of 14.23 which leads to rejection of the hypothesis at 5% significance level.

[30] The Likelihood Ratio test statistic is ....

[31] The OECD (1993) reports that this was the approximately the average tax rate paid by industrial workers in the United Kingdom over the period during which the dataset was

Given the estimated parameters in Table 2 we can evaluate equation (9) for each individual. These fitted values are the employment and unemployment rates after the tax has been imposed, as reported in the table 3. If we assume that the tax system is approximately proportional we can calculate the employment and unemployment rates before tax by simulating (9) using the gross wage in place of the net wage. Then using (10), we can calculate the monthly Compensating Variation per worker to be \$244 on average. This amount is sufficient to restore workers on average to their pre-tax utility given that they are now less likely to choose employment because of the imposition of the tax. The revenue raised by this tax is \$272 per worker per month on average. This figure must be adjusted for the fact that the imposition of the tax reduces the number of individuals working both directly as a result of the tax, but also indirectly as a result of the payment of the compensating variation.[32] This gives tax revenue of \$208 per month on average from each individual (as distinct from each worker).[33] Thus the DWL of the tax is \$35.89 per month or 26% of revenue.

The most striking feature of Table 3 is the size of the DWL. As reported in section 2, typical measures of the welfare cost of labour market taxes have been of the order of 2 to 3%. The estimates produced here are an order of magnitude greater. Furthermore, the pattern of the DWL across the income distribution is curious. The cost appears to be highest for those in the middle of the income distribution and lower for those at either end.

It is interesting to note that the DWL figures shown in Table 3 are very close to those reported by Hausman (1985) using the NLBS method. This fact helps shed some light on controversy over whether the NLBS method is biased. The existing literature on the estimation of labour supply models suggests that the treatment of the budget constraint matters. Pencaval (1986) reports that estimates of the compensated elasticity of supply given by the linearization method are smaller, by a factor of five, than the estimates using the NLBS procedure of Hausman (1985). McCurdy (1992) has argued that this difference is due to the fact that the Hausman method effectively imposes large compensated elasticities on the data. The results presented here suggest that the crucial issue may be the treatment of non-participants. The linearized method is typically estimated over the employed, so it returns the low compensated elasticity of hours conditional on participation. The NLBS method, being a generalized version of the Tobit procedure, easily accommodates the non-participants (but not the unemployed) generating an average of the participation and hours elasticities. Therefore the difference in measured elasticities could be entirely due to a high elasticity of participation.

The second two columns of Table 2 show estimates of the NMNL model. As can be seen, the coefficients are very similar to those estimated in the MNL model. With the exception of the coefficient on married in V^u, all the coefficients

collected.

[32] We follow Diamond and McFadden (1974) in calculating the tax revenue at the compensated point.

[33] Note that the revenue figure in the table is $E(\text{Wage} \times \text{tax} \times \text{prob. emp}) \neq E(\text{wage}) \times E(\text{tax}) \times E(\text{prob. emp.})$

in the NMNL model have the same sign as those in the MNL. Thus the basic results of the MNL model are confirmed by the NMNL model. A higher wage tends to increase the relative probability of being employed and reduce the relative probability of being unemployed. The coefficient on benefit income has a positive effect on the relative probability of being unemployed. As in the case of the MNL, the data cannot reject the hypothesis that wages have zero effect in the unemployment equation and benefits have a zero coefficient in the employment equation. Similarly, the model also rejects the hypothesis that the labour market can be modeled as a binomial choice (i.e. $\beta^e = \beta^u$ or $\beta^u = \beta^n$).

The coefficient on the inclusive value term, ρ, deserves special attention. McFadden (1986) notes that if the coefficient on the inclusive value is significantly different from unity, then the data rejects the hypothesis of independent errors necessary to estimate the model as a standard multinomial logit. As can be seen from Table 2, the data decisively rejects this hypothesis. We also know from section three that if $\rho > 1$ the model is not consistent with utility maximization. McFadden (1986) suggested that it may be that the tree structure implicit in the model is incorrect. Therefore we estimated an alternative tree structure was also estimated whereby U and N were hypothesized to be similar. This model produced nonsensical results (not reported here). In particular ρ turned out to be negative.

If the data and model are not consistent with utility maximization welfare calculations have little meaning. Nevertheless, Hausman et. al. (1997) report that empirical implementations of the NMNL model often produce estimates of ρ greater than unity. A justification for this is that the condition, $\rho < 1$ is sufficient for consistency with utility maximization, but it is not necessary. Therefore it is still worthwhile to report in table 4 the results for the DWL of tax using the NMNL estimates. This table was calculated in the same manner as table 3 using equations (11) and (14). Note that the DWL is slightly higher for the case of the NMNL model. The NMNL allows for a greater degree of similarity between alternatives than does the MNL. Therefore we might expect the welfare cost to be lower as there is less lost when individual moves from one alternative to another similar alternative. The fact that we do not observe this, is a further indication that the NMNL model may be misspecified.

If the NMNL model is misspecified then doubt is also cast on the validity of the MNL model. The MNL model is nested within the NMNL model, so if the latter is rejected so is the former. One obvious reason why these models could be misspecified is that unemployment is not, in fact, voluntary as was assumed. This leads to the estimation of the model of involuntary unemployment as discussed in section three.

Table 5 shows results for the joint (FIML) estimation of (17). The variables in employment probability equation are jointly and individually significant. This implies that occurrence of unemployment is not a random event but is correlated with variables that typically enter labour supply equations either directly or indirectly via the wage. This may not be surprising, but it is a factor that the traditional labour supply models tend to ignore. The individual estimated coefficients in this equation take on plausible values. The probability that the individual gets a job, conditional on his decision to participate, increases with increasing education. It falls with higher regional unemployment and is higher for married men than for single men. In addition it is higher for white men than for non-whites and ¯rst rises, then falls with increasing age.[34] The coe±cients of the indirect utility indices V^e and V^u are also mostly as expected. A higher wage increases the utility of employment and hence the probability of participation. Higher unearned income decreases utility of both employment and unemployment and hence the probability of participation. The only surprise is that higher bene¯ts appear to reduce the conditional utility of unemployment. This may indicate that bene¯ts facilitate non-participation, and so discourage job search.

As before we can calculate the welfare cost of taxation using these estimates. The calculations are reported in Table 6. This table was calculated in the same manner as table 3 using equations (17) and (18). The average DWL is of a similar order of magnitude as in the case of the two models of voluntary unemployment. The average welfare cost is actually lower in involuntary case. It is striking, however, the pattern across the wage distribution is so di®erent in this case. The distribution of the burden of taxes is heavily skewed towards low earners, with welfare cost falling (as a proportion of tax revenue) as the gross wage increases. We can get an idea of why this might be the case by looking at the empirical (unconditional) employment probability shown in Table 6. The participation rate is basically 100% for those individuals who would receive a wage in the highest quartile, if they secured employment. Thus the behaviour of these individuals is relatively insensitive to the tax. In the other two models, the tax could also distort the choice between employment and unemployment allowing the tax to have a greater a®ect on behaviour of the above the nedian wage.

6 Conclusions & Extensions

The goal of this paper was to calculate the dead-weight loss of taxes in a frame work that explicitly allowed for both unemployment and non-participation. Two di®erent classes of econometric models were applied to British panel data. Each model allowed individuals to be employed, unemployed or a non-participant.

The estimates con¯rmed some reasonable hypotheses regarding the labour market. Most particularly, it seems that employment, unemployment and non-participation are all qualitatively di®erent states. Any analysis of labour supply that fails to take this into account could be misleading.

The DWL ¯gures calculated from all the models are much larger than those typically estimated by labour supply models that adopt the linear approach to the budget constraint. For example, using the estimates of the traditional labour supply model reported in Table 1, I calculated the DWL of the 23% tax rate to be equal to 1:7% of revenue. The di®erence in size between this estimate

[34] The conditional probability of employment is highest when the respondant is 38 years of age.

and those using the discrete choice models, can be explained, to some extent, by the fact that the traditional estimators do not account for participation effects, whereas the multinomial models used in this paper do. Typically when compensated hours elasticities are estimated, on the assumption of a linear budget constraint, they turn out to be low (often less than 0.2).[35] Thus, conditional on participation workers are relatively insensitive to tax changes. The discrete choice models, effectively estimate participation elasticities, which are generally regarded as being larger than the standard hours elasticity.

The labour supply model and the models of section three are measuring different but related responses. The labour supply model measures the response to wages of hours worked conditional on participation and employment. This elasticity is small. Once they have made the decision to participate in the labour force and once they have a job, men are relatively insensitive to the wage. The discrete models, however, estimate participation elasticities. These elasticities are much higher. This is not surprising. If there are fixed costs of participation and job search then individuals' participation decisions will be very sensitive to the wage level, but once those fixed costs have been incurred, individuals behavior (hours supplied) is relatively insensitive to the wage level.

The calculated DWL of the same taxes will differ greatly depending on whether we focus on discrete choice or the hours model. The DWL of taxes based only on the hours model will be quite low, since workers are relatively insensitive to wage, the government can tax hours with relative impunity. On the other hand if we use the discrete models, we will get large dead weight losses, because the participation decision is very sensitive to the wage level.

Although not the original purpose of this paper, its results may help shed some light on the controversy regarding the alleged bias of the NLBS model method of Hausman (1985a). While, the DWL figures calculated here turned out to be much higher than those estimated by the traditional labour supply literature, they are close to those estimated using the NLBS method. This suggests that the higher elasticity estimated by the NLBS model could be due to fact that NLBS accounts for participation (if not unemployment) as here, whereas the linearized method typically does not.

This paper is a step towards the ultimate goal of a complete characterization of the impact of taxation on the labour market. Much remains to be done. In particular two important statistical extensions to this paper and one important theoretical extension suggest themselves.

Firstly, we must develop an estimation strategy that can incorporate variation in hours, without dropping the distinction between the three labour market states. Not accounting for the variation in hours means that we are throwing way information which could be used to identify individuals' preferences more accurately.

Another important extension to the statistical model is to allow for dynamics. It is quite likely that employment status is more persistent than the independent variables would suggest. This would occur, for example, if there

[35] See Pencaval (1986) or Blundell (1993) for surveys of the literature.

are large costs to both the employer and employee of breaking the match. This would tend to generate autocorrelated errors. In non-linear models autocorrelation leads to inconsistency as well as ine±ciency. If the persistence is due to the presence of some unobserved factor speci¯c to each individual, then the random e®ects estimators could control for it. More, generally, however, we would want to model the autoregressive nature of the errors directly i.e. the equivalent of including lags of the dependent variable as regressors in the estimated models.

The most important extension is to improve upon the partial equilibrium nature of the analysis in this paper. Speci¯cally we must explicitly model the interaction between the ine±ciency generated by taxation and the ine±ciency responsible for unemployment in the ¯rst place. This is important because the interaction could have very substantial implications for welfare. The imposition of even a small tax on an already distorted economy could have ¯rst order welfare e®ects. Alternatively, the two sources of ine±ciency could counteract each other, leading to a lower welfare costs of taxation than that reported here. Clearly the usual competitive general equilibrium models of tax incidence are unable to resolve this issue. Its resolution requires research into the theory of taxation in the labour market.

Lastly, the main result of this paper deserves restating. Even if the reader is sceptical regarding the appropriateness of the particular econometric speci¯cation used here, he must accept the basic result of this paper: how one chooses to model unemployment will signi¯cantly a®ect the estimates of the welfare cost of taxation of labour income.

References

[1] Ashenfelter, O. (1978). \What is Involuntary Unemployment". Proceedings of the American Philosophical Society, vol. 122, pp. 135-8.

[2] Atkinson, A. (1993). \Have Social-Security Bene¯ts Seriously Damaged Work Incentives in Britain?", in \Welfare and Work Incentives"ent Compensation

[3] Ben-Akiva, M. & Lerman, S. (1985) Discrete Choice Analysis: Theory and Applications to Travel Demand. MIT press, Cambridge, Massachusetts.

[4] Blomquist, N. (1995). \Restrictions in Labour Supply Estimation: Is the MaCurdy Critique Correct?". Economic Letters, vol. 47, pp. 229-235.

[5] Blundell, R. (1993). \UK Taxation and labour Supply Incentives", in \Welfare and Work Incentives"

[6] Chamberlin, G. (1980). \Analysis of Covariance with Qualitative Data". Review of Economic Studies. vol. 47, pp. 225-38.

[7] Diamond, P. and McFadden, D. (1974) \Some Uses of the Expenditure Function in Public Finance", Journal of Public Economics, vol. 3, pp. 3-21

[8] Flinn & Heckman (1983). \Are `Unemployed' and `Out of the Labour Force' Behaviourally distinct states". Journal of Labour Economics, vol. 1, pp. 28-42.

[9] Hausman, J. (1985). \Taxes and Labour Supply" in: Auerbach, A. and Feldstein, M., eds. Handbook of Public Economics, Vol I. (North Holland, Amsterdam).

[10] Hausman, J., Leonard G, & McFadden, D. (1997). \A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses due to Natural Resource Damage". Journal of Public Economics. vol. 56, pp. 1-30.

[11] Ham, J. (1982). \Estimation of a Labour Supply Model with Censoring due to Unemployment and Underemployment". Review of Economic Studies, vol. 49, pp. 335-354.

[12] Heckman, J. (1979). \Sample Selection as a Speci¯cation Error". Econometrica, vol. 47, pp. 153-61.

[13] Jones, S. and Riddell, W. (1999). \The Measurement of Unemployment: An Empirical Approach". Econometrica, vol. 67, pp. 147-162.

[14] Keane, M. (1992). \A Note on Identi¯cation in the Multinomial Probit Model". Journal of Business & Economic Statistics, vol. 10, pp. 193-200.

[15] Mroz, T. (1987). \The Sensitivity of Empirical Models of Married Women's Hours of Work to Economic and Statistical Assumptions". Econometrica, vol. 55, pp. 765-799.

[16] McCurdy, T. (1992). \Work disincentive e®ect of taxes: A re-examination of some of the evidence". American Economic Review, vol. 82, pp. 243-49.

[17] McFadden, D. (1986). \Econometric Models of Probalistic Choice", in \Structural Analysis of Discrete Data with Econometric Applications", Manski, C. & McFadden, D. (MIT press, Cambridge, Massachusetts), pp. 198-272.

[18] OECD (1986). Standardised Unemployment Rates: Quarterly Labour Force Statistics, Sources and Methods. OECD, Paris.

[19] OECD (1993). \The Tax and Bene¯t Position of the Average Production Worker". OECD, Paris.

[20] Pencaval, J. (1986). \Labor Supply of men". Handbook of Labour Economics. North-Holland.

[21] Small, K. and Rosen, H. (1981). \Applied Welfare Economics with Discrete Choice Models". Econometrica, vol. 49, pp. 105-131

Table 1: Hours Model

	OLS	2SLS
Net Wage (per hour)	-0.64	0.22
	(0.05)	(0.13)
Other income (per week)	-0.04	-0.08
	(0.02)	(0.01)
age	0.87	-0.02
	(0.01)	(0.02)
Kids	0.84	1.31
	(0.15)	(0.18)
married	4.21	-0.14
	(0.34)	(0.57)
constant	41.59	46.84
	(0.46)	(1.02)
Compensated Elasticity	0.13	0.41
DWL/R (%)	1.4	4.5

1. Standard Errors are in parentheses
2. Dep. Var. is usual hours worked per week

Table 2: Voluntary Unemployment

	MNL		NMNL	
	V^e	V^u	V^e	V^u
Net wage/1000	5.72	-	5.04	-
	(0.20)	-	(0.20)	-
Bene¯ts/1000	-	4.09	-	4.05
	-	(1.14)	-	(1.12)
Other income/1000	-4.96	{1.67	-3.72	-0.42
	(0.14)	(0.16)	(0.01)	(0.13)
Kids	-0.09	-0.03	-0.05	-0.02
	(0.09)	(0.10)	(0.06)	(0.08)
married	0.64	0.08	0.31	-0.35
	(0.16)	(0.16)	(0.12)	(0.14)
Constant	2.79	-2.43	4.12	-0.28
	(1.53)	(1.61)	(1.12)	(1.19)
Age	-0.17	0.24	-0.35	-0.01
	(0.13)	(0.13)	(0.09)	(0.09)
$Age^2/1000$	4.23	-3.75	7.93	1.46
	(3.08)	(3.26)	(3.27)	(2.57)
$Age^3/10^5$	-4.27	0.48	-6.17	-2.08
	(2.33)	(2.53)	(1.68)	(1.85)
Inclusive Value (1)	1		1.94	
	-		(0.14)	
Random E®ect	1.42		1.15	
	(0.12)		(0.08)	

1. Standard Errors are in parentheses

Table 3: DWL from MNL Model

	Total	Quartiles			
		(1)	(2)	(3)	(4)
Tax Rate	0.24	0.16	0.24	0.26	0.29
	(0.05)	(0.04)	(0.01)	(0.01)	(0.01)
Gross Wage ($)	1054	464	902	1217	1633
	(452)	(141)	(105)	(89)	(211)
DWL/R (%)	26.65	20.55	27.26	29.12	19.68
	(38.22)	(13.93)	(29.92)	(39.68)	(54.55)
DWL ($)	35.89	6.55	35.96	50.52	50.55
	(32.85)	(7.03)	(16.90)	(25.64)	(45.11)
CV ($)	244.28	41.01	171.11	298.82	466.17
	(169.75)	(37.59)	(63.14)	(58.58)	(86.87)
Revenue ($)	208.38	34.46	135.15	248.30	415.62
	(158.01)	(31.84)	(59.10)	(65.79)	(104.32)
Emp. rate after tax	0.75	0.43	0.72	0.88	0.96
	(0.30)	(0.32)	(0.26)	(0.17)	(0.09)
Unemp. rate after tax	0.13	0.22	0.17	0.08	0.03
	(0.12)	(0.11)	(0.12)	(0.08)	(0.04)
Emp. rate before tax	0.83	0.50	0.85	0.96	0.99
	(0.29)	(0.34)	(0.23)	(0.12)	(0.05)
Unemp. rate before tax	0.07	0.17	0.08	0.02	0.00
	(0.10)	(0.10)	(0.08)	(0.04)	(0.01)
Emp. Rate after tax and CV	0.67	0.40	0.62	0.77	0.87
	(0.28)	(0.29)	(0.24)	(0.18)	(0.14)
Unemp rate after tax and CV	0.18	0.23	0.24	0.17	0.09
	(0.12)	(0.12)	(0.12)	(0.09)	(0.07)

1. Standard Deviations are in parentheses
2. Quartiles are on the basis of the Gross Wage

Table 4: DWL from NMNL Model

	Total	Quartiles			
		(1)	(2)	(3)	(4)
Tax Rate	0.24 (0.05)	0.16 (0.04)	0.24 (0.01)	0.26 (0.01)	0.29 (0.01)
Gross Wage ($)	1054 (452)	464 (141)	902 (105)	1217 (89)	1633 (211)
DWL/R (%)	26.86 (40.68)	21.47 (16.47)	37.10 (34.78)	29.03 (46.01)	19.86 (53.23)
DWL ($)	35.46 (32.82)	6.56 (7.39)	34.65 (17.31)	48.71 (26.29)	51.93 (44.50)
CV ($)	244.57 (169.16)	41.42 (38.63)	172.53 (62.88)	298.83 (57.61)	465.51 (86.49)
Revenue ($)	209.11 (156.84)	34.86 (32.91)	137.88 (60.14)	250.11 (65.39)	413.57 (102.44)
Emp. rate after tax	0.75 (0.30)	0.44 (0.33)	0.73 (0.26)	0.89 (0.16)	0.96 (0.09)
Unemp. rate after tax	0.13 (0.12)	0.21 (0.13)	0.17 (0.13)	0.09 (0.08)	0.04 (0.05)
Emp. rate before tax	0.83 (0.29)	0.50 (0.35)	0.85 (0.22)	0.96 (0.11)	0.99 (0.05)
Unemp. rate before tax	0.07 (0.10)	0.17 (0.11)	0.09 (0.10)	0.03 (0.05)	0.01 (0.02)
Emp. Rate after tax and CV	0.67 (0.28)	0.41 (0.31)	0.64 (0.25)	0.78 (0.18)	0.87 (0.13)
Unemp rate after tax and CV	0.19 (0.12)	0.22 (0.13)	0.24 (0.13)	0.18 (0.10)	0.11 (0.08)

1. Standard Deviations are in parentheses
2. Quartiles are on the basis of the Gross Wage

	V^e	V^u	Employment Probability
Net wage/1000	1.99	-	-
	(0.43)	-	-
Bene¯ts/1000	-	-7.11	-
	-	(5.22)	-
Other income/1000	-2.01	-0.47	-
	(0.11)	(0.27)	-
Kids	-0.12	0.68	-
	(0.11)	(0.77)	-
married	0.64	-0.98	0.36
	(0.19)	(0.36)	(0.03)
constant	-0.65	-1.42	-1.32
	(0.87)	(1.95)	(0.16)
Age	0.12	0.11	0.14
	(0.05)	(0.12)	(0.01)
Age^2/1000	-1.67	-1.92	-1.79
	(0.58)	(1.46)	(0.07)
U. rate/100	-	-	-5.17
	-	-	(0.59)
White	-	-	0.24
	-	-	(0.08)
educ1	-	-	0.42
	-	-	(0.07)
educ2	-	-	0.40
	-	-	(0.04)
educ3	-	-	0.46
	-	-	(0.04)
educ4	-	-	0.63
	-	-	(0.06)
educ5	-	-	0.72
	-	-	(0.07)
educ6	-	-	0.76
	-	-	(0.14)
Random E®ect	1.11		
	(0.06)		

1. Standard Errors are in parentheses

Table 6: DWL from Nested Probit Model

	Total	Quartiles			
		(1)	(2)	(3)	(4)
Tax Rate	0.24 (0.05)	0.16 (0.04)	0.24 (0.01)	0.26 (0.01)	0.29 (0.01)
Gross Wage ($)	1054 (452)	464 (141)	902 (105)	1217 (89)	1633 (211)
DWL/R (%)	22.82 (24.72)	39.44 (16.07)	29.68 (28.70)	14.75 (22.71)	7.42 (14.87)
DWL ($)	26.70 (19.19)	12.08 (9.48)	34.74 (14.69)	32.74 (18.50)	27.26 (22.59)
CV ($)	257.46 (165.49)	51.22 (42.82)	193.16 (55.95)	313.43 (45.92)	472.02 (81.52)
Revenue ($)	230.75 (161.94)	39.15 (34.75)	158.42 (55.13)	280.69 (53.16)	444.76 (87.14)
Emp. rate after tax	0.76 (0.27)	0.46 (0.31)	0.75 (0.20)	0.89 (0.11)	0.95 (0.05)
Unemp. rate after tax	0.12 (0.09)	0.19 (0.11)	0.16 (0.07)	0.09 (0.04)	0.05 (0.02)
Emp. rate before tax	0.77 (0.26)	0.47 (0.31)	0.77 (0.18)	0.89 (0.09)	0.95 (0.04)
Unemp. rate before tax	0.13 (0.09)	0.20 (0.11)	0.17 (0.07)	0.09 (0.04)	0.05 (0.02)
Emp. Rate after tax and CV	0.75 (0.27)	0.45 (0.31)	0.73 (0.21)	0.87 (0.12)	0.94 (0.07)
Unemp rate after tax and CV	0.12 (0.09)	0.19 (0.11)	0.15 (0.07)	0.09 (0.04)	0.05 (0.02)

1. Standard Deviations are in parentheses
2. Quartiles are on the basis of the Gross Wage

Table 7: Summary of DWL Measures

	Total	Quartiles			
		(1)	(2)	(3)	(4)
Tax Rate	0.24	0.16	0.24	0.26	0.29
	(0.05)	(0.04)	(0.01)	(0.01)	(0.01)
MNL	0.27	0.21	0.37	0.29	0.20
	(0.38)	(0.14)	(0.30)	(0.40)	(0.55)
NMNL	0.27	0.21	0.37	0.29	0.20
	(0.41)	(0.16)	(0.35)	(0.46)	(0.53)
Nested Probit	0.23	0.39	0.30	0.15	0.07
	(0.25)	(0.16)	(0.29)	(0.23)	(0.15)

1. Standard Deviations are in parentheses
2. Quartiles are on the basis of the Gross Wage

Table 8: BHPS Data

Variable	De¯nition	Mean	Stn. Dev.
age	age at date of interview	40.98	14.43
health	=1 if su±ciently healthy for work	0.85	0.35
marstat	marital status	-	-
kids	number of own children in house	0.55	0.95
region	region / metrapolitan area	-	
pasoc	father's occupation	-	-
tae	age left school/college	17.77	4.21
educ	higher value{higher ed	-	
white	=1 if white	0.97	0.18
other income	reported monthly non-labour income	228.91	371.14
k02y	no. of children aged 0-2	0.08	0.28
k34y	no. of children aged 3-4	0.08	0.28
k511y	no. of children aged 5-11	0.25	0.6
k1215y	no. of children aged 12-15	0.17	0.45
k1618y	no. of children aged 16-18	0.04	0.21
nwage	observed monthly net wage		
gwage	observed monthly gross wage		
ben	observed monthly bene¯ts		
net wage	predicted monthly net wage	781.39	300.08
gross wage	predicted monthly gross wage	1054.04	451.66
bene¯ts	predicted monthly bene¯ts	225.41	77.55
U. rate	regional male unemployment rate	12.48	2.41
married	=1 if married	0.64	0.48
hoh	=1 if head of household	0.75	0.43

1. Statistics are calculated for the pooled cross section

	Selection for wage		Net Wage		Gross Wage		Bene¯ts		Selection for Bene¯ts	
Dep. Var	nwage>0		ln(nwage)		ln(gwage)		ln(ben)		ben>0	
Indep. Vars.	Coef.	t-stat.	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
cons	-3.62	-9.98	4.07	22.23	3.90	18.83	3.91	8.95	-1.05	-3.00
age	0.22	8.58	0.17	14.71	0.20	15.17	0.08	2.85	0.02	1.09
$age^2/10^3$	-3.29	-5.46	-3.26	-11.63	-3.81	-12.01	-1.13	-1.99	-0.57	-1.29
$age^3/10^4$	0.05	1.04	0.21	8.71	0.25	9.04	0.00	0.25	0.04	1.44
k02y	-0.38	-6.69	0.10	4.74	0.12	4.86	0.26	3.86	0.38	6.31
k34y	-0.30	-5.13	0.06	2.72	0.07	2.85	0.26	3.68	0.34	5.58
k511y	-0.20	-7.14	0.02	2.00	0.02	1.89	0.11	3.12	0.17	5.66
k1215y	-0.09	-2.40	-0.01	-0.75	-0.02	-1.13	0.17	3.49	0.17	4.14
k1618y	-0.15	-2.10	-0.19	-6.70	-0.20	-6.41	-0.01	-0.12	0.06	0.65
pasoc1	-0.10	-0.81	0.17	3.44	0.21	3.75	-0.24	-1.26	-0.30	-1.99
pasoc2	0.15	1.35	0.15	3.37	0.18	3.66	-0.13	-0.81	-0.26	-2.02
pasoc3	0.08	0.68	0.07	1.42	0.08	1.62	-0.07	-0.43	-0.05	-0.33
pasoc4	0.07	0.69	0.07	1.72	0.09	1.93	-0.11	-0.68	-0.01	-0.10
pasoc5	0.01	0.13	0.09	2.09	0.12	2.29	0.02	0.10	0.13	0.96
pasoc6	0.03	0.22	0.04	0.71	0.05	0.89	-0.06	-0.36	0.15	1.04
marstat2	-0.85	-7.91	0.15	3.08	0.20	3.51	-0.05	-0.43	0.97	9.29
marstat3	-0.68	-10.29	0.03	1.02	0.07	2.06	0.01	0.13	0.85	12.69
marstat4	-0.24	-1.86	-0.35	-5.07	-0.34	-4.38	0.16	1.44	0.45	4.90
marstat5	-0.55	-10.46	-0.01	-0.23	0.03	1.11	-0.19	-2.58	0.43	7.44
white	0.31	4.06	-0.02	-0.69	-0.02	-0.55	-0.20	-2.08	-0.31	-3.64
health	1.14	27.34	-0.23	-5.09	-0.27	-5.33	-0.09	-1.68	-0.30	-6.68
hoh	-0.05	-0.98	0.13	6.97	0.16	7.68	0.06	1.01	-0.14	-2.57
educ1	-	-	0.16	6.18	0.19	6.40	-	-	-	-
educ2	-	-	0.17	10.56	0.21	11.19	-	-	-	-
educ3	-	-	0.23	13.54	0.27	14.02	-	-	-	-
educ4	-	-	0.34	14.60	0.40	15.12	-	-	-	-
educ5	-	-	0.42	19.86	0.49	20.30	-	-	-	-
educ6	-	-	0.49	14.01	0.55	14.04	-	-	-	-
region2	-	-	-	-	-	-	-	-	-0.22	-1.31
region3	0.35	3.14	-	-	-	-	-	-	-0.31	-2.46
region4	0.15	1.10	-	-	-	-	-	-	-0.26	-1.78
region5	0.22	1.21	-	-	-	-	-	-	-0.13	-0.65
region6	0.39	2.72	-	-	-	-	-	-	-0.46	-2.83
region7	0.13	0.74	-	-	-	-	-	-	-0.13	-0.67
region8	0.26	1.56	-	-	-	-	-	-	-0.34	-1.72
region9	0.36	1.97	-	-	-	-	-	-	-0.18	-0.94
region10	-0.47	-2.11	-	-	-	-	-	-	0.30	1.33
region11	-0.12	-0.72	-	-	-	-	-	-	0.02	0.13
region12	0.01	0.04	-	-	-	-	-	-	0.05	0.24
region13	0.38	1.98	-	-	-	-	-	-	-0.40	-1.85
region14	0.33	1.54	-	-	-	-	-	-	-0.27	-1.14
region15	0.16	0.74	-	-	-	-	-	-	-0.44	-1.79
region16	0.10	0.56	-	-	-	-	-	-	-0.26	-1.33
region17	0.14	0.87	-	-	-	-	-	-	-0.37	-2.03
region18	0.16	1.20	-	-	-	-	-	-	-0.45	-2.94
region2*time	-0.89	-0.19	-	-	-	-	-	-	-2.49	-0.43
region3*time	0.12	0.04	-	-	-	-	-	-	-4.48	-1.19
region4*time	7.58	1.73	-	-	-	-	-	-	-1.10	-0.23
region5*time	-7.26	-1.12	-	-	-	-	-	-	2.19	0.32
region6*time	-7.71	-1.69	-	-	-	-	-	-	4.30	0.81
region7*time	-0.57	-0.09	-	-	-	-	-	-	-1.02	-0.15
region8*time	-0.28	-0.05	-	-	-	-	-	-	-3.95	-0.55
region9*time	-5.52	-0.84	-	-	-	-	-	-	-0.55	-0.08
region10*time	18.64	2.12	-	-	-	-	-	-	-19.57	-2.05
region11*time	8.12	1.35	-	-	-	-	-	-	-11.28	-1.63
region12*time	-4.70	-0.64	-	-	-	-	-	-	-4.29	-0.54
region13*time	-13.75	-2.02	-	-	-	-	-	-	6.75	0.86
region14*time	0.31	0.04	-	-	-	-	-	-	0.14	0.02
region15*time	-6.73	-0.83	-	-	-	-	-	-	10.96	1.25
region16*time	1.63	0.26	-	-	-	-	-	-	0.05	0.01
region17*time	-3.74	-0.71	-	-	-	-	-	-	4.19	0.69
region18*time	-1.88	-0.44	-	-	-	-	-	-	5.64	1.14
Mills ratio			-0.67	-9.76	-0.78	-10.11	0.3	6.02	-	-
R^2/pseudo R^2	0.42		0.42		0.43		0.81		0.65	