

## **Evaluating Labour Market Interventions.**

**Donal O'Neill<sup>1</sup>**

NUI Maynooth, March 2000

*Abstract:* The high growth rates experienced in Ireland over the last 10 years has resulted in a tightening of the labour market which is reflected in the number of unfilled vacancies reported by firms. At the same time wage inequality has increased leading to greater demands being placed on the government to tackle social exclusion. In response to these issues, recent governments have proposed a range of policies involving direct intervention in the labour market. Effective implementation of these policies requires careful monitoring and evaluation of their effects. In this paper I examine the procedures currently available for evaluating labour market interventions. The results of recent evaluations of minimum wages laws, reform of the benefit system and changes in working-time conditions are used to illustrate the methodologies involved. I also describe the data requirements of these methodologies and examine the currently available Irish labour market data in this light.

---

<sup>1</sup> I would like to thank Gerry Boyle, Denis Conniffe, Sarah Feehan, John Kennan, Frances Ruane and Olive Sweetman and an anonymous referee for comments on earlier versions of this paper.

## ***1. Introduction.***

Between 1990 and 1997 real GDP in Ireland increased by over 6% per-annum on average. The corresponding figure for the European Union as a whole was less than 2%. One consequence of these rapid growth rates has been a significant tightening of the labour market. This is reflected in the number of job vacancies currently on offer and the reported difficulties that firms are having in filling these openings.<sup>2</sup> A potential worry is that these pressures will fuel increased wage demands, which if not matched by increased productivity may reduce the competitiveness of Irish firms. At the same time however, wage inequality among workers has increased (Barrett et al 1997) and the share of output going to capital, as opposed to labour, has also risen substantially (Lane 1998). These trends have resulted in greater demands being placed on the government to deal with social exclusion.

In response to both these concerns, recent governments have proposed a number of policies that involve direct intervention in the labour market. These include; the proposed national minimum wage; reform of the unemployment benefit system with tighter controls being placed on those receiving unemployment benefits in return for greater assistance in obtaining work; and restrictions on the working week. While these policies may be well intended the predictions of economic theory concerning their likely effects is often ambiguous.

Monopsony models of the labour market can predict higher employment levels after small increases in the minimum wage. A competitive model on the other hand would predict job losses as a result of introducing a binding minimum wage. Furthermore these job losses would be concentrated among unskilled workers, who are precisely the individuals the policy was supposed to help<sup>3</sup>. With regard to welfare reform we would expect that greater job search assistance would ease the transition from welfare to work. However in doing so it may also lead to increased wage pressure in the labour market which could in turn have a negative effect on employment rates (Calmfors 1994). Alternative models of the labour market can also lead to ambiguities when considering the effects of working time restrictions. Imposing restrictions on the standard working week may increase employment levels if more workers are required to produce the same level of output. However by altering the composition of the wage bill such a policy may lead to a substitution of hours for

---

<sup>2</sup> For a more detailed discussion of vacancies in Ireland see Williams and Hughes (1999).

<sup>3</sup> For a discussion of the economics of the minimum wage see Dolado et al (1996).

employees, a substitution of capital for labour and a scale effect which would reduce the level of output produced. All of these effects would result in a fall in employment as a result of the reduced working week. (Freeman 1998).

It is clear from this that the effective implementation of these policies requires careful monitoring and evaluation of their labour market effects. The need for ex-post evaluations is also reflected in the requirements of the Structural Fund regulations, which require that Community structural operations be subject to appraisal, monitoring and ex-post evaluation. However, conducting such evaluations is not straightforward. In this paper I examine the procedures currently available for evaluating labour market interventions. These include regression-based procedures; procedures relying on data from controlled experiments and procedures which make use of natural experiments. Results of recent evaluations of minimum wage laws, reforms of the benefit system and reductions in the standard working week are used to illustrate the procedures.

By highlighting the ambiguities in the theory and emphasizing the need for empirical evaluation I do not wish to give the impression that such evaluations should be atheoretical. Indeed, quite the opposite is true. Throughout this paper it will become evident that researchers who ignore available theory will have difficulty in choosing the appropriate estimation technique and in constructing the appropriate data on which to apply their estimator. Furthermore, they will encounter problems interpreting their estimates.

In section 2 of the paper I outline the basic problem facing researchers attempting to evaluate the impact of labour market interventions, namely the problem of selection bias. Section 3 considers three alternative estimation procedures proposed to overcome this problem. I illustrate each with examples from recent research. Section 4 discusses the data requirements of these procedures and analyses current Irish labour market data in this light and section 5 summarizes the main results of the paper and offers some conclusions.

## ***2. Sample-Selection Problem***

To understand the problems that may arise with evaluations consider a situation in which there are two potential states: a treated state in which the intervention is in place and an untreated state in which the intervention is absent. Let  $Y_T$  denote the potential outcome if the treatment is received and  $Y_C$  the potential

outcome when it is not.<sup>4</sup> The impact of the intervention on person  $i$  in period  $t$ ,  $\Delta_{it}$ , is given by:

$$\Delta_{it} = Y_{itT} - Y_{itC} \quad (1)$$

It is clearly not possible to calculate  $\Delta_{it}$  for each individual. If an individual receives the treatment we observe  $Y_{iT}$  but cannot observe  $Y_{iC}$ . Alternatively the individual does not receive the treatment, in which case we observe  $Y_{iC}$  but cannot observe  $Y_{iT}$ . Formally, the observed outcome,  $Y_i$ , is given by

$$Y_i = D_i * Y_{iT} + (1 - D_i) * Y_{iC}$$

where,  $D_i$  is an indicator variable, taking the value 1 if the individual receives the treatment and zero otherwise. At any given time therefore we observe only one of two mutually exclusive possibilities.

The same is true if we focus on the average gain or loss in the population  $E(\Delta)$ . While we can estimate  $E(Y_T)$  for participants using  $(\overline{Y_T})_1$ , and  $E(Y_C)$  using  $(\overline{Y_C})_0$  (where the subscripts outside the brackets denote participation status), we cannot observe either  $E(Y_C)$  for participants or  $E(Y_T)$  for non-participants.

The goal of the estimators discussed in this paper is to allow the researcher to construct these unobserved counterfactuals. If the treatment is universal one might use the outcomes prior to the implementation of the program (say at time  $t'$ ) to proxy for what would have occurred in time  $t$  if the program had not been implemented. In this case  $\{(\overline{Y_{tT}}) = (\overline{Y_t})\}$  and  $\{(\overline{Y_{tC}}) = (\overline{Y_{t'}})\}$ . Formally the *before-after* estimator can be defined as:

$$\hat{\Delta} = \overline{Y_t} - \overline{Y_{t'}} \quad (2)$$

This estimator can be rewritten as :

$$\hat{\Delta} = (\overline{Y_t} - \overline{Y_{tC}}) + (\overline{Y_{tC}} - \overline{Y_{t'}}) \quad (3)$$

The first term in the expression measures the parameter of interest. Unless  $(\overline{Y_{tC}} - \overline{Y_{t'}}) = 0$  the before-after estimator of this parameter will be biased. This latter term represents the selection bias for this estimator and arises because of differences between the true and proxied counterfactual outcomes. If governments find it easier to

---

<sup>4</sup> This section of the paper follows closely the exposition provided in Heckman, LaLonde and Smith (1999).

implement legislation during economic recoveries we might expect  $(\overline{Y}_{IC} - \overline{Y}_t) > 0$ , in which case the before-after estimator would overestimate the impact of the legislation.

We can illustrate this using figure 1. Since treatment is universal we can ignore the line labeled control for the moment. In keeping with our earlier conjecture we see that the prior to the introduction of the legislation the outcome variable (Y) had been increasing for several periods. In the absence of the treatment we would have expected output to reach E by the time of the evaluation (in our earlier notation E represents  $\overline{Y}_{IC}$ ). As a result of the treatment we observe the outcome B ( $\overline{Y}_t$ ). The impact of the program is (B-E). The before-after estimator uses outcomes prior to the introduction of the legislation (for example A ( $\overline{Y}_t$ )) to proxy for the unobserved counterfactual. In this instance the before-after estimator would be given by (B-A) > (B-E). The distance (E-A) measures the selection bias.

If the program is voluntary we might consider using non-participants to proxy for the unobserved outcomes. In this case we might estimate the average gain for participants in the program as :

$$\hat{\Delta} = (\overline{Y}_T)_1 - (\overline{Y}_C)_0 \quad (4)$$

where the subscripts “1” and “0” outside the brackets again denote participation and non-participation respectively. This cross-section estimator can be rewritten as :

$$\hat{\Delta} = [(\overline{Y}_T)_1 - (\overline{Y}_C)_1] + [(\overline{Y}_C)_1 - (\overline{Y}_C)_0] \quad (5)$$

The first term in this expression is the average impact of the treatment on the treated. The second term represents the selection bias for this estimator. Again the bias arises because of differences between the average outcome of non-participants and the outcome we would have observed for participants if they had not participated. If more motivated individuals participate in training, for example, we would expect this latter term to be positive. In this case comparing participants and non-participants would overestimate the impact of the program. Looking at figure 1 we see that participants (labeled treatment) have uniformly higher outcomes than non-participants (labeled control) even prior to the legislation. The cross section estimator would estimate the average effect of the treatment on the treated as (B-D) which again in this model is greater than the true effect (B-E). In this model the selection bias is measured by (E-D).

In the next section of the paper I will discuss alternatives to the simple before-after and cross-section estimator, illustrating each with examples from current research. In doing so it will become clear that all of these estimators are based on identifying assumptions, many of which may be difficult to test. In order to choose between competing estimators it may be necessary to turn to theories explaining the underlying process.

### ***3. Identification Strategies for Evaluating Labour Market Interventions.***

#### ***3.1 Social Experimentation.***

A key to estimating the impact of a program is the construction of plausible comparison groups. Social experiments solve this problem by creating a random control group of individuals who would have, under normal circumstances, been affected by the treatment but who were prevented from receiving the treatment by the evaluators. Recent examples of the use of social experiments to evaluate labour market interventions can be found in Card and Robins (1998), Ashenfelter et al (1998), Dolton and O'Neill (1995, 1996a, 1996b, 1999), Ham and LaLonde (1996), Gorter and Kalb (1996), Johnson and Klepinger (1994), Davidson and Woodbury (1993) and Decker (1994). Earlier examples include Burtless and Hausman (1978), LaLonde (1986), Corson and Nicholson (1985), Corson et al (1989) and Johannesson and Persson-Tanimura (1978).<sup>5</sup> Surveys of this approach can be found in Burtless (1995), Meyer (1995) and Heckman, LaLonde and Smith – henceforth HLS (1999).

To see how social experiments work, consider a program in which participation is universal. Let  $(Y^r_i)_1$  and  $(Y^r_i)_0$  denote outcomes in the presence of random assignment of participants and non-participants respectively. The social experiment estimator of the average treatment effect is given by :

$$\hat{\Delta}_r = \overline{(Y^r)_1} - \overline{(Y^r)_0} \quad (6)$$

This estimator identifies the treatment effect provided that:

$$E[(Y^r)_1 - (Y^r)_0] = E(Y_T - Y_C) \quad (7)$$

---

<sup>5</sup> The results of this latter study are summarised in Robinson (1995).

This assumption rules out systematic changes resulting from the process of randomization itself.<sup>6</sup> If assumption (7) holds the average treatment effect can be estimated by comparing the randomized control and treatment groups.

To illustrate this approach consider the recent study by Dolton and O'Neill (1999) who use experimental data to analyze the impact of the Restart unemployment program in the U.K.. The Restart program consists of a compulsory interview for the long-term unemployed with an official of the Employment Office. The aim of this interview was to reduce individual welfare dependency. The Restart process combines positive help and encouragement given to the unemployed job seekers by way of advice, counseling and direct contact with employers, with tighter enforcement of the conditions necessary to qualify for UB.

We examined the impact of Restart by comparing unemployment outcomes for a treatment group with those of a randomly chosen control group for whom participation in the process was postponed for six months.<sup>7</sup> An important feature of the Restart experiment was the ability to match the experimental data with the Joint Unemployment and Vacancies Operating System (JUVOS) data collected by the Employment Service. This allowed us to construct monthly work histories for each individual in the sample over a 15 year period, as a result of which we could track individuals for 5 years after the end of the experiment.

The availability of data prior to the commencement of the experiment also allowed us to examine the validity of the randomization procedure. Average unemployment rates for both the treatment and control groups are presented in figure 2. Randomization took place in April 1989 (indicated by the first vertical line on figure 2). To test the validity of the random assignment we compare unemployment rates prior to April 1989. A comparison of the unemployment rates prior to 1989 reveals no significant differences between the two groups.

Under assumption (7) we can estimate the impact of Restart by comparing average unemployment rates for the treatment and control groups after April 1989.<sup>8</sup> To focus on the long-run effects of Restart we compare average unemployment rates in

---

<sup>6</sup> An example of randomization biases arises when the process of experimentation alters individual behaviour. Such affects are commonly called Hawthorne effects. For a discussion of the original Hawthorne experiments see Franke and Kaul (1978).

<sup>7</sup> The randomization procedure was based on National Insurance numbers.

<sup>8</sup> Since members of the control group received a Restart interview if they were continuously unemployed for 1 year the Restart experiment actually measures the impact of postponing search assistance and work search tests by 6 months.

1994, 5 years after the original experiment. By this stage the unemployment rate for individuals excluded from the initial Restart interview was 6 percentage points higher than those receiving the treatment. We used these estimates to conduct a cost benefit analysis of the Restart program. This showed that the introduction of the scheme resulted in significant savings for the government. These results are consistent with the conclusions of others (Calmfors (1994), Meyer (1995a) and HLS (1999)) all of whom suggest that counseling and job search assistance, targeted at the long-term unemployed, provide a cost effective way of increasing job-finding rates. Our findings indicate the impact of these programs persists in the long-run. Previous concerns that these programmes simply initiate a circular flow of individuals through the benefit system<sup>9</sup>, are not supported in our data.

These findings have implications for Active Labor Market Policies in Ireland (ALMP). A recent OECD study (OECD (1996)) ranks 14 countries based on spending on ALMP. The table distinguishes between different types of spending, including spending on job search assistance, training and subsidies. Of the 14 countries Ireland ranked third in total spending on ALMP, with total spending amounting to approximately 1.3% of GDP (Sweden (2.18%) and Denmark (1.86%) were the top two). However when we look at spending on job search assistance Ireland was only ranked ninth out of the 14 countries. Our findings indicate that a redistribution of existing funding may be a more efficient use of the funds.<sup>10</sup> Calmfors (1994) reaches a similar conclusion for a broad range of countries:

“Counseling activities and job search assistance, should probably be given a greater weight in such a [ALMP] portfolio, since a fair amount of evidence seems to indicate a favourable impact on job-finding rates, and serious adverse side effects appear unlikely.” (page 37).

The experimental approach to program evaluation has been the subject of substantial debate and some of the limitations of the approach have been discussed in

---

<sup>9</sup> See for example Robinson (1995) and Disney et al (1992).

<sup>10</sup> In recent years the Irish government have introduced changes to the Irish benefit system, which incorporate elements of the Restart scheme. The most notable of these has been the Employment Action Plan, which commenced on September 1<sup>st</sup> 1998. From that date persons aged under 25 who reach six months on the live-register are referred by the Department of Social Community and Family Affairs for interview with FAS. To date of the 26,0000 called for interview, approximately 70% are no longer signing on the Live Register. While the threat component of the scheme seems to be having an effect, it is unfortunate that the counseling and search assistance component of Restart does not seem to have been as well developed.

recent studies (Heckman and Smith (1995), HLS (1999) and Blundell (1999)). In analyzing experimental evaluations it is important to realise that other estimators share some of the problems associated with experiments. Blundell (1999) cites the expensiveness of experiments as a drawback. It is true that experiments tend to be rare and can be expensive to implement. However, a large proportion of this cost does not reflect the cost of randomization but rather the cost of data collection (HLS (1999)). A major problem with non-experimental evaluations of ALMP is that often the constructed comparison groups do a poor job of proxying the required counterfactual. HLS argue that the best solution to this problem is to obtain better data. They point out that obtaining this data may prove to be almost as expensive as conducting an experimental evaluation. Thus when comparing estimators of similar quality, it is not clear, a priori, which will be cheapest.

Two other features of social experiments have been criticised (Heckman and Smith 1995). Firstly there is the possibility that treatment group members drop out of the program before receiving the treatment and secondly there is the possibility that some control group members may receive treatments which are close substitutes for the experimental treatment (substitution bias). Again, however, neither of these problems is unique to social experiments. There is no evidence that the rate of dropping out is higher in an experimental setting than a non-experimental setting.<sup>11</sup> It has been suggested that experimentation increases the likelihood of substitution bias by creating a pool of individuals who expressed a desire to receive the treatment but were randomized out of the treatment in the experiment. However this is not likely to be much of a problem for evaluations of new treatments, such as the Restart experiment. In the Restart experiment almost all of the treatment group recalled getting the letter inviting them to the interview (the first stage of the Restart process). Almost 80% are recorded as having attended a Restart interview and less than 2% of the control group are recorded as receiving a Restart interview.

In accessing social experiments it is also important to distinguish between problems arising from the experimental process itself and those unique to specific experiments, arising from poor design. Heckman and Smith (1995) note that 90% of training centres refused to participate in the experimental evaluations of the Job Training

---

<sup>11</sup> Even if subjects drop out, Heckman, Smith and Taber (1998) have shown that experimental data can still be used to estimate the impact of the treatment on the treated under plausible assumptions provided one can observe the proportion of dropouts.

Partnership Act (JTPA) in the U.S.. This can lead to geographical biases in the final estimates and may also contaminate the randomization procedure if eligibility requirements for the program are relaxed to ensure a sufficiently large sample. However we found little evidence of such problems in the Restart experiment. Every Employment service office throughout Britain was contacted while constructing the sample and there was no scope for local employment offices to opt out. By only having a small proportion of the total office caseload participate in the experiment (about 1 in 100 of those interviewed over a period) the Restart design also minimized the extra burden placed on staff as a result of the experiment.

Keeping the proportion of participants low in each centre also reduces the potential for training officials to manipulate the services in favor of the treatment group (the ‘creaming’ effect). In an ideal design, officials would not know the identity of those taking part in the experiment. In the Restart experiment, however, counselors knew the identity of those taking part in the experiment. Nevertheless we are confident that there was little scope for creaming. White and Lakey (1992) note that because the participants accounted for such a small proportion of the total number of clients, this meant that pressures of the caseload left little scope for Restart counselors to give special treatment to those in the experimental sample.

While some of the problems associated with experiments are shared by other approaches and some can be overcome by careful design of the experiment there are some limitations that are specific to the experimental approach, which are difficult to overcome by appropriate redesign of the experiment.

While the experimental approach may be suitable for analyzing the average treatment effect in a population it is not suited to answering questions about the distribution of these gains. A training program that increase average wages by 5% may result from a situation where every participant gains by 5% or one where half of the participants gain by 15% and half experience a fall in their wages of 5%. Social experiments cannot distinguish between these two possibilities. The problem arises because average impacts may be estimated using the marginal distribution of outcomes for the treated and untreated. These can be obtained from social experiments. However a distributional analysis requires knowledge of the joint distribution of outcomes in the treated and untreated state, which is not provided by the experimental approach.

Heckman and Smith (1995) and Heckman, Smith and Clements (1997) discuss work by Frechet and Hoeffding that demonstrates how to bound individual elements of a

frequency table from knowledge of the marginal distributions. To examine the width of these bounds I use the Restart data for 1994, 5 years after the completion of the experiment. The proportion of the treatment group employed at this stage was .692, while the proportion of control group members employed was .632. However the estimated bounds on the proportion of employed workers who moved out of employment as a result of the scheme is [.06,.368]. Unfortunately these bounds are quite wide which means we cannot distinguish between a situation where Restart helps many people back into work but also harms the employment prospects of others, and one where Restart helps fewer individuals but harms nobody. The former seems unlikely given the nature of the Restart program. For training programs however, there is some evidence of locking-in effects (Calmfors 1994), whereby individual search intensities may be reduced during the periods of programme participation with detrimental effects on employment prospects. For policymakers interested in the distribution of these gains, experimental evaluation procedures may be of limited use.

It may also be difficult to use social experiments to analyse training programs that involve several distinct components. If the program is successful it is often difficult to ascertain what component of the program was most valuable (Meyer 1995a).<sup>12</sup> In theory it is possible to overcome this problem by distinguishing between different components in the randomization process. In practice, however, this is not always straightforward. In the Washington experiment analysed by Johnson and Klepinger (1994) different combinations of eligibility checks and job search assistance were assigned to different treatment groups. However, many individuals received the letter inviting them to the jobsearch workshop prior to their scheduled Eligibility Review interview. Even though these individuals were to be excluded from the eligibility check, it seems as though many of them perceived the jobsearch invitation as a further work-test and reacted accordingly. This may have biased the cross-component comparisons of the experiment.

Finally social experiments provide no information on the process governing individuals decisions to participate in the scheme. It has often been noted that adult unemployed males are less likely to participate in training programs than younger males or females. Even if experiments provide an unbiased estimate of the effect of the

---

<sup>12</sup> This is somewhat true of the Restart experiment discussed above. However, Dolton and O'Neill (1999) attempt to assess the relative importance of the work-test and job-search assistance components of Restart.

program on participants they provide no information on why adult males fail to enter the program or how best to alter the program to encourage participation. These are important questions and ones which policy makers may be interested in answering.<sup>13</sup>

Given the advantages and limitations of social experiments what role should they play in program evaluation ? It is my view that carefully designed experiments provide a straightforward way of obtaining clean estimates of the average treatment effect. They are best suited to analysing new projects, as opposed to ongoing projects and the stage of the process at which randomization takes place may have important consequences for what can be learned. In designing experiments, data should be collected from a large number of centres, keeping the proportion of participants in each centre small. While this may increase the administrative costs of conducting the experiments, it has the advantage of limiting the incentives for creaming and at the same time increasing the probability of participation for a given centre. However, even a well designed experiment may have little to say about distribution or participation effects. Policymakers interested in these issues will have to supplement the experimental evidence with alternative estimates.

### 3.2 Difference in Difference Estimator (DDE) and 'Natural' Experiments

An alternative to carrying out controlled randomized experiments, is to look for examples of experiment-like situations occurring naturally. This is the logic behind the 'natural experiment' literature and the 'difference-in-difference' estimator that is commonly applied in these circumstances. To see how the estimator works assume that there exists a control group, denoted by 0. Also assume we have data before and after the intervention. Retaining the notation from section 2, let  $\hat{\Delta}_1 = (\bar{Y}_t - \bar{Y}_{t'})_1$  denote the before-after difference in outcomes for those who receive the treatment and  $\hat{\Delta}_0 = (\bar{Y}_t - \bar{Y}_{t'})_0$  denote the average difference in outcomes for the comparison group. The difference in difference estimator is defined as :

$$\hat{\Delta}_{DD} = (\hat{\Delta}_1 - \hat{\Delta}_0)$$

---

<sup>13</sup> Although it is difficult to analyse participation issues using the experimental approach one recent study by Card, Robins and Lin (1997) uses the approach to evaluate whether targeted incentives to leave welfare can increase participation welfare.

The key identifying assumption underlying this estimator is that any differences between the control and treatment groups, in the absence of the legislation, is time-invariant. This implies that the average *change* in outcomes, *in the absence of the intervention*, is the same for participants and non-participants.<sup>14</sup> We can illustrate the DDE using figure 1. The distance (B-E) gives the true impact of the legislation. The first part of the difference in difference estimator is simply the before-after estimator for the treatment group (B-A). The line AE denotes what would have happened to the treatment group in the absence of the legislation. To obtain the true treatment effect we would need to subtract (E-A) from (B-A). However, under the identifying assumption of the estimator  $(E-A)=(D-C)$ . Thus the treatment effect can be estimated by  $(B-A)-(D-C)$ .

Examples of studies using the DDE include Card (1990), Eissa and Liebman (1996), Katz (1996) and Hamermesh and Trejo (2000), while Meyer (1995b) provides a detailed discussion of natural experiments and the difference-in-difference estimator. However, the best known recent study may be the work of Card and Krueger on minimum wages (Card and Krueger (1994)). Their treatment group comprised of fast-food restaurants in the state of New Jersey, while the control group comprised of similar establishments in the neighbouring state of Pennsylvania. In April 1992, New Jersey's minimum wage rose from \$4.25 to \$5.05, while legislators in the state of Pennsylvania kept the minimum wage constant. Card and Krueger use the DDE to compare employment, wages and prices at stores in New Jersey and Pennsylvania before and after the minimum wage. They concluded that increases in the minimum wage had only a negligible effect on employment.<sup>15</sup>

For these estimates to be reliable it is essential that the treatment and control groups satisfy the identifying conditions required for the estimator. Changes in macroeconomic conditions may not affect all groups in the same way and it is crucial to bear this in mind when implementing the procedure. Situations where treatment group members anticipate the intervention and act accordingly or where the intervention takes effect only with a significant lag may also cause problems. Furthermore not every law change is a good natural experiment and simply labeling a source of variation a 'natural experiment' does not ensure that the variation is exogenous.

---

<sup>14</sup> In some cases the difference in difference estimator may be extended to allow for the possibility of time-varying affects – the so-called difference-in-difference-in-difference estimator (Meyer 1995).

<sup>15</sup> These studies have formed an important component of a larger debate on the overall consequences of minimum wages on employment (see for e.g., Neumark and Wascher (1997) and the response by Card and Krueger (1997))

To illustrate these points I consider a recent change to the legislation governing the standard working week in Canada.<sup>16</sup> In 1993 the province of Newfoundland introduced a 40-hour standard working week for all employees. Any hours worked in excess of the standard working week must be paid and overtime premium. The standard came into force on December 31, 1993. As far as we can establish the motivation behind this legislation seems to have been the removal of industry-wide discrepancies in the standard work week, rather than an attempt to increase employment. The neighbouring province of Nova Scotia already had a 40 hour standard working week for all employees which was not changed.

These circumstances seem to provide an ideal opportunity to study the effects of reductions in the standard working week. It seems reasonable to treat the change in the standard working week as exogenous. The change in hours was pronounced (from 44 to 40 hours). The retail sector in Newfoundland was already subjected to the 40 working week and thus could potentially be used to control for time invariant differences across states. Finally aggregate data on employment, hours and wages on a monthly basis were readily available.

The data for full-time workers in both states are presented in figure 3. It would be difficult to reject the DDE identifying assumption based on these data. Although there are differences between the two states these differences appear to be constant. The 'difference-in-difference' estimates for these data indicate that the legislative change reduced employment by 1,106 workers, corresponding to an hours' elasticity of  $-.06$ . This very small elasticity is consistent with recent work on hours legislation (Hunt 1999) and suggests that contrary to much popular belief such laws are unlikely to have a significant effect on employment levels.

The flaw in this experimental design only becomes apparent when one analyses the micro data on hours worked available from Statistics Canada. Figures 4a and 4b plot hours worked per week in Newfoundland before and after the legislative change. In *both* cases we see a significant spike in the hours distribution at 40 hours. It appears that even prior to the official change in the standard week many workers in Newfoundland were already working a 40 hour week. What our original estimate shows, therefore is that a law that does not affect anyone has no effect.

---

<sup>16</sup> This works draws heavily on research by Feehan (1999).

We can also use employment levels within the construction industry in both states to further illustrate the potential difficulties with DDE. These data are plotted in figure 5. As before the pre-legislation data would suggest that the identifying conditions needed for the DDE are satisfied. For this sector the `difference-in-difference' estimate of the employment effect implies an elasticity of over +1. This estimate implies that reductions in the standard working week have the potential to significantly increase employment. However, we already know that the legislation affected very few people, so this can not be so. It must be the case that the hours legislation is proxying for some omitted factor which had differential effects on employment in Newfoundland and Nova Scotia.

This example provides a clear illustration of some of the pitfalls associated with the DDE and highlights a central theme of Meyer's (1995) survey. While natural experiments can provide a straightforward procedure for evaluating labour market interventions with modest data requirements, the researchers cannot experimentally control the variation being used in these studies. In this case it is essential that they understand the source of this variation. The preceding example shows that this is not always straightforward.

### 3.3 Regression Based Approaches

While both randomized and natural experiments are being used more frequently in recent evaluations, the most common approach to estimating labour market interventions is based on traditional regression approaches. There exist several studies summarizing these approaches (Maddala (1993), Friedlander, Greenberg and Robins (1997), Vella (1998)). In this section I provide a brief overview of the regression procedures available, emphasizing the relationship between these approaches and the estimators discussed earlier. These procedures centre on the estimation of a system of equations describing the intervention. We can illustrate these approaches by examining training schemes. If participation in a training scheme is voluntary and the returns to training are homogeneous then the training process can be described by the following system of equations<sup>17</sup> :

$$Y_i = \alpha + \beta X_i + \delta T_i + u_i \quad (8.1)$$

---

<sup>17</sup> In the evaluation literature this model is often called the common coefficient model or the endogenous dummy variable model (Heckman 1978).

$$T_i^* = \gamma W_i + e_i \quad (8.2)$$

$$\begin{aligned} T_i &= 1 \quad \text{if } T_i^* > 0 \\ T_i &= 0 \quad \text{if } T_i^* < 0 \end{aligned} \quad (8.3)$$

The first equation describes the outcome for individual  $i$ . The remaining equations describe the training participation decision.  $T^*$  is a latent variable describing training propensity and  $T_i$  reflects the binary training decision. Selection bias arises here because of correlation between  $T_i$  and  $u_i$ . As a result the ordinary least squares (OLS) estimates of  $\delta$  are biased. This arises because OLS mistakenly attributes variation in  $Y$  resulting from variation in  $u_i$  to the regressor,  $T_i$ .

There are two channels through which this correlation may arise. Firstly  $W_i$  and  $u_i$  may be correlated. This type of selection model is called selection on observables. Alternatively  $e_i$  and  $u_i$  may be correlated, in which case we have selection on unobservables. There have been a number of estimation procedures developed to deal with these situations. When selection is based on observables consistent estimates of  $\delta$  may be obtained by including the  $W$  variables in the outcome regression. It can be shown that rather than including all of the  $W$  regressors in the outcome equation, it is sufficient to include only a particular function of the  $W_i$ , called the propensity score. This is simply the probability of treatment conditional on  $W_i$ . For an application of the propensity score method to the evaluation of training programmes see Dehejia and Wahba (1999).<sup>18</sup>

When selection is based on unobservables the proposed solutions are more complex. One popular estimation strategy used in this instance is the instrumental variable approach. The instrumental variable approach relies on obtaining a variable ( $Z_i$ ) which is correlated with  $T_i$  but uncorrelated with the error term in the outcome equation,  $e_i$ . In carrying out the estimation only the variation in  $T_i$  explained by  $Z_i$  is used to produce the estimate. The instrumental variables approach is currently the centre of an ongoing debate.<sup>19</sup> Heckman (1997) shows that the conditions necessary for consistent IV estimation of the average treatment effect are quite restrictive when the returns to training vary across individuals. Verifying these conditions is difficult and requires detailed knowledge of the participation process (Heckman and Robb 1985). When these conditions are not satisfied the IV procedure estimates a weighted average of the

---

<sup>18</sup> For a more detailed discussion of the propensity score method see Conniffe et al (2000).

<sup>19</sup> See for example the discussion in Angrist, Imbens and Rubin (1996a, 1996b) and Heckman (1996).

individual returns where the weight is proportional to the causal effect of the instrument. Instruments that focus on particular groups, therefore estimate the marginal return for this specific group, which may be higher or lower than the average treatment effect.

A common alternative to the IV approach is the two-stage procedure developed by Heckman (1979). Heckman showed that this problem could be reformulated within the standard omitted variable framework. In this case the omitted variable is the mean of  $u_i$ , conditional on individual  $i$ 's participation decision. Heckman's proposed solution was to estimate the variable giving rise to the specification error (the conditional error terms) and to include this in the outcome equation. Estimates of the conditional error terms are constructed from the estimated parameters of the participation equation (equation 8.2).<sup>20</sup>

Estimation of the participation decision is also of interest in its own right and may provide answers to questions of concern to policy makers. As noted in Heckman and Smith (1995) policymakers may be interested in the effects of factors such as subsidies, local labour markets, family income, race age and sex on program participation decisions. The estimated participation equation based on observational data provides a framework for addressing these issues. This is in contrast to data traditionally collected in randomized experiments, which, by the nature of their design, are often not capable of addressing these issues.

The conventional regression approach can also be modified to allow the impact of the treatment to differ across participants. If we assume that the gain to treatment  $\delta_i$  is independent of the base state,  $Y_{iC}$ , we can formulate the model as a random coefficient regression model.<sup>21</sup> When estimated this model allows us to retrieve the distribution of gains across individuals. A significant advantage of this approach, therefore, is that it allows one to address the distributional aspects missing from the social experimental studies.

While estimation of structural models eliminate the need for experimental data, this is achieved by restricting the range of plausible models. The two stage approach outlined above requires an additive separable model relating regressors to outcomes and also an additive error structure. The model postulated in 8.1-8.3 assumes that we have a continuous outcome variable (such as earnings) and that the censoring appears in the

---

<sup>20</sup> This model can also be modified to allow for heterogeneity in rewards, which in turn allows for the estimation of more parameters than the simple average treatment effects (see for example Bjorklund and Moffitt (1987)).

<sup>21</sup> For a discussion of random coefficient models see Swamy et al (1988a, 1998b, 1989).

selection equation. Consider a model however where censoring occurs in both the primary and the selection equation. We can write this system as follows :

$$Y_i = \beta X_i + \delta T_i + u_i \quad (9.1)$$

$$T_i^* = \gamma W_i + e_i \quad (9.2)$$

$$Y_i = 1 \text{ if } Y_i^* > 0, Y_i = 0 \text{ otherwise} \quad (9.3)$$

$$T_i = 1 \text{ if } T_i^* > 0, T_i = 0 \text{ otherwise.} \quad (9.4)$$

The second stage (equation 9.2) is similar to equation 8.2. However the first stage differs, in that the primary equation is censored rather than continuous. We only observe whether an individual is employed or not ( $Y_i$ ), which is related to an individual's employability,  $Y_i^*$ .  $Y_i^*$  however is unobserved. In the absence of selection bias both these models can be estimated using traditional methods, such as a probit or logit model (Amemiya 1981). If the errors are correlated however this procedure is not valid. The model could be estimated using maximum likelihood procedures if a joint distribution for the error terms is specified. As it stands however, it is not valid to estimate this model using the traditional two-stage approach. The problem is that including the Heckman correction in the primary equation alters the error distribution of that equation, thus invalidating the use of a second stage probit<sup>22</sup>. Many of these traditional regression approaches to program evaluation are available in statistical packages such as Stata or Limdep. As a result there is a tendency to apply these procedures without giving sufficient thought to the structure of the underlying model or data. The example provided above illustrates the problems this can cause.

In a recent study Heckman et al (1998) showed that the traditional bias in program evaluations could be decomposed into three sources. The first source of bias occurs when it is not possible to match individuals in the treatment group with individuals in the control group with similar characteristics. The second source of bias arises if, for those individuals that can be matched, the *distribution* of X differs between the two groups. The final type of bias captures selection on unobservables.

The relative magnitude of each of these components of the bias can be informative in deciding on which estimator is best suited to the evaluation. Combining experimental and observational data, Heckman et al show that although the bias resulting

---

<sup>22</sup> For a further discussion of models, which are not suited to the traditional two step approach, see Maddala (1983) and Vella (1999).

from selection on unobservables is large relative to the estimated treatment effect, it is relatively small compared to other components of the bias. They conclude that better data for modeling participation decisions and outcome variables can help a lot in evaluating ALMP. Unfortunately these findings do not establish whether better data can help for broader types of interventions. Would controlling for aggregate measures of demand and supply eliminate the potential bias in time series studies of the minimum wage? If not would other data help and where might these data be found? In the remainder of this paper I outline some of the key data sets available for studying Irish labour market interventions. I also highlight some issues, which if tackled, would further facilitate the evaluation of labour market interventions in Ireland.

#### ***4. Data***

In describing the available data I distinguish between primary data sets which are collected with a specific evaluation in mind and secondary data sets which are collected for purposes other than program evaluation but which often prove useful in conducting evaluations. I begin with the latter type of data.<sup>23</sup>

##### *4.1 Secondary Data Sets.*

The largest publicly available household data sets in Ireland are the Household Budget Surveys (HBS) and the Labour Force Surveys (LFS). The main purpose of the HBS is to generate weights that can be used in constructing price indices. However the micro data also contain information on employment, income and expenditure. To date micro data sets for 1987 and 1994 have been made available by the Central Statistics Office. These contain information on 7,705 and 7,877 nationally representative households respectively. Examples of studies using the HBS include Nolan and Callan (1994) and O'Neill and Sweetman (1999), Van de gaer, Funnel and McCarthy (1999) who use the HBS to study inequality in Ireland, McCarthy (1998) who studies tax reform and Madden (1999) who examines the measurement of poverty.

The HBS is a valuable source of information on labour market and social issues and the public release of the micro data represented an important development in the analysis of labour market phenomena in Ireland. However, there are a number of

---

<sup>23</sup> In this section I will focus on the usefulness of current data for evaluations of labour market interventions. However the availability of micro data plays an important role in evolution of labour economics in general (Rosen 1990). The current and continuing availability of micro data is to be welcomed and will help develop our understanding of many labour market issues, including the determinants of labour supply, skill shortages, wage differentials, inequality and unemployment.

features that make it less suitable for studying specific labour market interventions. At present the HBS is only carried out every seven years, only information on the head of household have been made available<sup>24</sup> and only very aggregate information regional information is available. There are plans to conduct future surveys on a 5 year cycle. While this move is to be welcomed, the 5 year window between surveys will still constrain potential users of the data. In contrast, the U.K. Family Expenditure Survey (FES) is a continuous survey that has been in existence since 1957. Individual micro data are available from the U.K data archive at Essex to all bona fide researchers. For a recent study that uses the FES to analyse labour market issues see Harmon and Walker (1995). Furthermore the HBS data are cross-section and contain only limited retrospective data on labour market issues. As noted by HLS “[when evaluating ALMP] it is important to ... draw the treatment and comparisons groups from the same local labour markets. In addition, recent evidence suggests that labour force status dynamics represents an important determinant of participation in job training programs.” In this context the absence of detailed regional data and the lack of a dynamic component in the HBS data restricts their potential for evaluating labour market interventions.

The Labour Force Surveys<sup>25</sup> consisted of a nationally representative survey of approximately 47,000 households undertaken in April/May each year. The LFS was official data source for the construction of unemployment statistics in Ireland. LFS microdata files are available from the CSO from 1994 to 1997. These data sets have been used to analyse unemployment duration, non-employment and unemployment dynamics in Ireland (see Murphy and Walsh (1996, 1997) and to construct comparison groups for the evaluation of government training schemes (Deloitte & Touche and Murphy (1998)).

This latter study illustrates some of the problems facing researchers conducting labour market evaluations on currently available data. The aim of the Deloitte & Touche and Murphy (DTM) report was to evaluate the impact of the Community Employment (CE) scheme. The CE scheme provides temporary, part-time, employment for the long-term unemployed and the socially excluded. It is the largest single programme operated by FAS – Training and Employment Authority. Employment records for CE participants

---

<sup>24</sup> In the past staff at the CSO have agreed to run specific programs on individual data. However this places an extra burden on the CSO staff and often several runs may be needed before a researcher decides on an appropriate specification.

<sup>25</sup> The Labour Force Survey was replaced by the Quarterly National Household Survey in 1997. I will discuss these data in more detail later.

were obtained from the 1996 and 1997 FAS Follow-up Surveys which were conducted by the ESRI. To construct a control group, data from the 1993, 1996 and 1997 LFS surveys were used. Given the structure of the LFS this is not straightforward. Individuals in the FAS follow-up survey would have entered the FAS scheme in 1994 and thus would have been unemployed about 20 months prior to the date at which employment information is observed. From the 1996 LFS data one can identify individuals who were unemployed 12 months earlier, which form the basis of their control group. Comparing members of the control group with CE participants indicated that the employment rates for male CE participants were at least 6.5 percentage points higher than for similar unemployed males. However, the properties of the constructed control group are open to question. Because of the cross-section nature of the LFS data DTM were forced to introduce a correction for state dependence - the fact that exit rate from unemployment decline with time spent unemployed – to estimate the employment rates for control group members who were usually unemployed 20 months earlier. Unfortunately, our understanding of the extent and nature of state dependence is still limited and so any allowances for it must necessarily be somewhat ad hoc. One may also question the extent to which the limited characteristics available for construction of the control group account for all possible selection biases. Finally DTM acknowledge that because training status is not coded in the LFS it is possible that some members of the control group may have been on FAS schemes. This makes it difficult to interpret their final estimates.

In September 1997 the Annual Labour Force Survey was replaced by the first in a series of Quarterly National Household Surveys. Information is collected continuously throughout the year, with 3,000 households surveyed per week giving a total of 39,000 households each quarter. Households are asked to take part in the survey for five consecutive quarters and then are replaced by other households drawn from the same local area.

The introduction of the QNHS represents a significant development in the collection of labour market data in Ireland. While there have been some changes to the questionnaire,<sup>26</sup> the content of the QNHS is for the most part similar to LFS. However, in a changing labour market more frequent information is needed and the QNHS gives a comprehensive picture of the labour market four times a year. The ability to match individuals across quarters also provides researchers with a basis for performing more

---

<sup>26</sup> The main effect of these changes has been to elicit more information on part-time employment (reference is 21<sup>st</sup> July 1999 CSO Press Release).

dynamic evaluations.<sup>27</sup> Another important development in the QNHS is the possibility of including supplemental questionnaires with the main survey. The 1998 June/August quarter had a module on housing, while the September-November quarter had a module on crime and victimisation. Survey supplements have been successfully included in the Current Population Surveys in the U.S and have been the source of some interesting studies of labour market issues (see for example (Vroman 1991)). If these supplemental questionnaires are to fulfill their potential it is important that researchers are involved in developing the modules. In this light it is encouraging to see that the review of the National Statistics Board (1998) notes that “.. a process of wide consultation will be used to ensure that these social modules provide the most valuable data possible.” While consultation with predesignated groups of experts is important, other forms of involvement should be considered. A call asking researchers to submit proposals suggesting ideas for supplemental questionnaires, to be considered on a competitive basis and subject to cost considerations, would mark an exciting development in research in labour economics in Ireland. Such provisions currently exist in other countries. Since 1993 The Panel Study of Incomes and Dynamics (PSID) in the U.S. has held open competitions among researchers to add supplemental questions to the PSID.

In contrast to the household-based surveys described above the unit of analysis in both the Census of Industrial Production and Forfas panel survey is the firm. The Census of Industrial Production provides a detailed picture of the structure and activity of industry over time. Short-term indicators of trends in industrial production, turnover, employment and earnings are published regularly to monitor current developments. Unfortunately these data cannot be accessed outside the CSO. The Forfas data are publicly available and have been used by Walsh, Strobl and Barry (1998) to analyse job flows in Irish economy. While these data contain useful information on firms very little detailed information is provided on employees.

As well as these data sets, there exist other Irish data sets containing important labour market information. The most notable of these are the Living in Ireland Surveys. The Living in Ireland Surveys comprises the Irish element of the European Community Household Panel and are conducted by the ESRI. The surveys are annual, the first taking place in 1994 and contain detailed socio-economic information on a representative panel

---

<sup>27</sup> The availability of matched data also provides researchers with access to a range of estimators based on panel data that are not discussed in this paper. For a discussion of these estimators see Vella (1999).

of individuals.<sup>28</sup> Recent work by O'Connell (1999) uses the LIS survey to carry out a control group analyses of 14 training and temporary employment schemes implemented by FAS. The main finding of this part of this study is that of the 14 schemes examined, 5 seem to have a significant impact on the probability of reemployment. These include the Employment Incentive Scheme, Employment Subsidy scheme, Enterprise Allowance, Specific Skills Training and Job Training programmes. On the other hand the Community Employment scheme appears to have no affect on reemployment prospects.<sup>29</sup> O'Connell uses the LIS to construct the comparison group. Unlike the LFS, the LIS is a panel data set that follows individuals over time. Thus for each individual in the data set there is a longitudinal record of their work histories. There is no need, therefore, for the type of state dependence corrections used by DTM. However, even then the control group constructed on the basis of the LIS is not entirely satisfactory, a point readily acknowledged by the author. Members of the control group are substantially older than the participants, significantly more likely to be married, less highly educated and have spent more time out of the labour force. These differences are accounted for in the study by controlling for variables such as age, marital status and education in the outcome equation. However if such differences exist between the two groups on the basis of observables then we should not be surprised if similar differences were also present in terms of unobservables. To allow for this O'Connell uses a two-stage Heckman approach. However as we noted earlier this approach restricts the range of plausible models and requires identifying restrictions which are often difficult to find. The difficulties with this approach are apparent in that controlling for selection bias reduces the employment effects by a margin of perhaps one-third, yet tests for selection effects do not reject the hypothesis of no selection bias. This may indicate problems with the identifying restrictions.

A problem with the LIS is that the data are not easily accessible. The European Statistical Office (Eurostat) coordinates the production and distribution of the ECHP. ECHP data access conditions and prices (Marlier (1999)) are prohibitive and have been criticised in recent work by Jenkins (1998). If these data are to be used to their full

---

<sup>28</sup> For a general description of the research carried out by the ESRI using these data see *ESRI* (1999).

<sup>29</sup> This finding seems to differ from that reported by Deloitte & Touche and Murphy. However given the different approaches adopted in these studies one comparing the results requires some work. We can get a rough idea of the differences by noting that O'Connell's study (Table 4) predicts the probability of a single male aged 25-39 with no qualifications and unemployed less than 6 months in 1994 who participated in a CE scheme was 30%. For a similar individual not on a CE scheme the probability is

potential it is vital that universal access to micro data be provided to all bona fide non-commercial users (subject to registration and license agreements) at a nominal cost.<sup>30</sup> This may involve a greater role for national data archives in distributing data (Jenkins 1998). The establishment of the National Social Science Data Archive, which is a joint venture between UCD and the ESRI, is a welcome development in this respect.

#### 4.2 Primary Data Sets

In some situations it may not be possible to use existing data sets to carry out the required evaluation. In these cases it might be necessary to collect new data. Primary data sets are becoming more common in the study of labour market issues and have been used in studies of homelessness (Freeman and Hall 1986), disadvantage (Freeman 1990), education (Ashenfelter and Krueger 1994) and the low wage labour market (Greene, Machin and Manning 1996), Card and Krueger (1994), Neumark and Wascher (1997).

The data sets used by Card and Krueger and Neumark and Wascher have been the subject of many debates<sup>31</sup> which have tended to focus on issues concerning the merits of personal versus telephone interviews, strategic response and biased sampling. However, an important lesson to be learned from the minimum wage studies is the need to recognize the underlying economic theories when collecting data. Much of the debate surrounding the minimum wage studies could be dispensed with (or at the very least replaced by a more constructive form of debate) if these data had included information that would allow us to understand how the labour market works. Recent work by Nolan, O'Neill and Williams (1999) represents an attempt to collect such information. The focus of their study was the national minimum wage being of £4.40 which was introduced in Ireland on April 1<sup>st</sup> 2000. In late 1998 Nolan, O'Neill and Williams (1999) undertook a survey of private establishments to obtain details on employment structure by hourly pay rates and the extent of vacancies, hirings and departures.<sup>32</sup> To avoid potential biases all questionnaires were completed on a personally administered basis, which involved an interviewer paying a visit to each respondent and completing the survey on site.

---

27%. The average estimates across all individuals in the DTM study (Table 6.21) are 30.6% and 23.7% respectively. The fact that DTM do not report standard errors further complicates the comparison.

<sup>30</sup> For a related discussion see Wagner (1998).

<sup>31</sup> See for example the papers by Neumark and Wascher (1997) and Card and Krueger (1997).

<sup>32</sup> For a more technical discussion of the survey design see Nolan, O'Neill and Williams (1999).

The information from the survey of firms was used to estimate the numbers of workers affected by a national minimum wage of £4.40. Our estimates indicate that approximately 21% of all private sector employees received an hourly wage of £4.50 or less. This is consistent with estimates obtained from household based surveys (Nolan (1999)). We believe the survey will also play an important role in the future monitoring and evaluation of the impact of the minimum wage. We plan to interview these firms again after the minimum wage is introduced, which will allow us to determine how they have been affected by the legislation

Earlier I argued that primary surveys should be designed so that the information obtained can be used to improve our understanding of the labour market. This requires an understanding of the economics of the problem. A key feature of recent monopsony models of the labour market is the rigidities that may exist in the labour market. Search frictions and turnover costs are often used to justify these rigidities. With this in mind, our survey included a range of questions on vacancies, hirings and departures of workers within establishments by pay range. Including these questions will allow us to establish the extent to which these rigidities exist for the type of workers affected by the minimum wage. They will also allow us to examine the role these factors play in determining the impact of a minimum wage. It is only by understanding the labour market processes underlying both the hiring decisions of firms and labour supply decisions of workers that we can satisfactorily make sense of the employment effects of previous studies.

### 4.3 Future Developments ?

#### *a. Linked Employer-Employee Database*

The existing data sets described in section 4.2 tend to focus almost exclusively on individuals or on establishments. However the outcome of labour market interventions often depends on the interaction between employers and workers. In these cases, data sets that exclude either one or other of these groups can only provide partial and possibly misleading answer to the problem being studied. One solution is to develop linked employer-employee data sets (Hamermesh (1990, 1999), Leonard (1999), Manser (1997). Having a data set with broad detailed questions on the nature and structure of firms, combined with accurate detailed information on the workers within these firms creates the potential for developing greater insight into the operation of the labour market.

An example of a matched employee-employer database is the 1998 U.K. Workplace Employee-Employer Relations Survey. The WERS 98 Panel Survey dataset contains data from interviews conducted with management respondents at the same establishment in both 1990 and 1998. Furthermore self-completion questionnaires were distributed to a random selection of up to 25 employees in each workplace. The majority of these data are available from the Data Archive at the University of Essex. A bibliography of research based on each of these surveys is available from the WERS98 Data Dissemination Service web-site<sup>33</sup>

Linked employer-employee data sets would also facilitate a greater understanding of the impact of policies such as the minimum wage or changes in standard hours. The impact of these policies is determined by the interaction of employers and workers and is best understood by a data set which reflecting this. The prospect of matching a survey of firms, such as the one described in Nolan, O'Neill and Williams (1999), with detailed individual information would significantly enhance our understanding of the labour market.

While the benefits of linked data sets are evident, such an effort could only be justified if a framework could be developed which would allow improved access to confidential micro establishment data. The sponsors of WERS98 achieve this by withholding a very small fraction of the data collected in separate data files. Access to these files is restricted, until a time to be agreed. These files are only released with the explicit permission of the sponsors. Providing access to the information contained in the restricted files could lead to the identification of individual workplaces and individual respondents. So as to guard against this, the WERS sponsors have decided that users must: a). give a considered account of why they need information contained in the restricted files over and above that contained in the general release; and, b). provide details on the steps that will be followed to preserve the anonymity of the data. These restricted files contain data on the regional location of the establishment and its detailed industrial activity (a broad, one-digit industrial classification is available on the general release files) With the exception of these data the remainder of the WERS data are available through the data archive.<sup>34</sup>

---

<sup>33</sup> The address for this website is <http://www.niesr.ac.uk/niesr/wers98/>. This website also contains downloadable versions of the questionnaires used in the survey.

*b. Live Register Data Panel*

While interventions such as minimum wage legislation are likely to have their largest effect on employees other legislation, such as welfare reform, concentrates directly on the unemployed. To date studies evaluating such changes have tended to rely on once off special surveys<sup>35</sup> or samples drawn from nationally representative samples. A problem with household surveys is that the samples of unemployed tend to be quite small and they provide little information on dynamics. These problems can be overcome by establishing a panel database focusing on welfare dynamics. The Joint Unemployment and Vacancies Operating System (JUVOS) in the U.K could serve as a benchmark.

The JUVOS cohort is a 5% sample of all computerised claims for unemployment related benefits selected by reference to a claimant's National Insurance Number. Each time a person with a relevant National Insurance Number makes a claim for unemployment-related benefits their details are added to the cohort file. The data are used to inform policy decisions on employment and training, welfare and social security. The availability of JUVOS was crucial in establishing the long-run impact of Restart discussed in section 3.

A limitation of JUVOS is the lack of information available on individuals. One can establish the age, gender and region of residence of the individual but no information is available on education, training, family structure, or the destination of individuals who leave the unemployment register. Individuals who leave the register could have exited to a job, a training scheme or out of the labour force. Knowing which, could have important policy implications. Therefore, to allow such a database to reach its potential it is important that future data developments should, where possible, facilitate linkages to any proposed live register panel. This would allow researchers to supplement JUVOS information with more detailed information from external sources. I understand that there have been interdepartmental discussions about implementing a longitudinal data set of this nature and any such moves should be encouraged.

## **5. Conclusion**

---

<sup>34</sup> For a discussion of possible data access provisions being discussed within the U.S Bureau of Labor Statistics see Manser (1997).

<sup>35</sup> See for example O'Connell and McGinity (1997).

In this paper I examined some of the procedures available for analysing labour market interventions and discussed the currently available data in this context. The paper outlines the identifying assumptions underlying each of the procedures and illustrates some of the issues involved with analyses of welfare reform, reductions in standard working hours and minimum wage changes. In evaluating these techniques, the tendency has been to view them as competing procedures. A theme that emerges from this paper, however, is that often these approaches can be complementary and when used in tandem can be more effective than the sum of their parts.

This is evident in the discussion of the distributional issues involved in program evaluation. Well-designed experiments provide sharp estimates of the average treatment effects but are relatively uninformative about the distributional consequences of the program. However, when combined with models of program participation the distributional bounds associated with social experiments can be reduced substantially. Likewise there exist many conventional regression approaches for dealing with sample selection, including matching models, instrumental variables and two stage approaches. Which approach is more appropriate depends on the nature of the selection bias. As we discussed in section 3 recent studies have shown how randomized data can be combined with observational data to estimate the relative importance of various forms of selection bias.

Careful evaluation of labour market interventions is essential if programmes are to work effectively. Conducting empirical evaluations requires choosing an identification strategy suited to the problem at hand. There is no one estimator that is applicable in all circumstances. Choosing between alternative (groups of) estimator(s) is one important element in developing an appropriate strategy. Ultimately, this will be determined by the questions to be answered, the economics of the problem and the availability of appropriate data.

### *References*

- Amemiya, T, (1981), "Qualitative Response Models: A Survey," *Journal of Economic Literature*, Vol. 19, no. 4, pp. 481-536.
- Angrist, J, G. Imbens and D. Rubin, (1996a), "Identification of Causal Effects Using Instrumental Variables" *Journal of the American Statistical Association* , Vol. 91, pp. 444-455.
- Angrist, J, G. Imbens and D. Rubin, (1996b), "Rejoinder to Heckman et al " *Journal of the American Statistical Association* , Vol. 91, pp. 468-472.
- Ashenfelter, O. and A. Krueger (1994), "Estimates of the Economic Return to Schooling from a new Sample of Twins," *American Economic Review*, Vol 84, no. 5, pp. 1157-73
- Ashenfelter, O, D. Ashmore and O. Deschenes (1998), "Do Unemployment Insurance Recipients Actively seek work ? Randomised Trials in Four U.S. States," Princeton University Working Paper #412.
- Barrett, A, T.Callan and B.Nolan (1997), "The Earnings Distribution and the Returns to Education in Ireland," ESRI research paper.
- Bjorklund, A. and R. Moffitt (1987), "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, vol. 69, no. 1, pp. 42-49.
- Blundell, R. (1999), "Welfare Reform and Low Income Workers," mimeo University College London.
- Burtless, G and J. Hausman (1978), "The Effects of Taxation on Labour Supply: Evaluating the Gary NIT Experiment," *Journal of Political Economy*, vol 86, no., 6, pp. 1103-30.
- Burtless, G (1995), "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, Vol. 9, no. 2, pp. 63-84.
- Callan, B. and B. Nolan (1994), Poverty and Policy in Ireland, Gill and Macmillan Ltd., Dublin.
- Calmfors, L (1994), "Active Labour Market Policy and the Unemployment – A Framework for the Analysis of Crucial Design Features," *OECD Economic Studies*, No. 22, Spring, pp. 7-47.
- Card, D, P. Robins and W. Lin (1997), "Would Financial Incentives for Leaving Welfare Lead Some People to Stay on Welfare Longer ? An Experimental Evaluation of `Entry Effects' in the self-sufficiency Project," Princeton University Working Paper #380.
- Card, D. and P. Robins (1998), "Do Financial Incentives Encourage Welfare Recipients to Work ?," *Research in Labor Economics*, 17, pp. 1-56.
- Card, D. (1990), " The Impact of the Mariel Boatlift on the Miami Labor Market, " *Industrial and Labor Relations Review*, vol. 43, pp. 245-57.

Card, D. and A. Krueger (1995), Myth and Measurement: The New Economics of the Minimum Wage, Princeton University Press, Princeton, New Jersey.

Card, D. and A. Krueger (1994), "Minimum Wages and Employment : A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, vol. 84, pp. 772-93.

Card, D. and A. Krueger (1997), "A Reanalysis of the Effect of the New Jersey Minimum Wage Increase on the Fast Food industry with Representative Data," Princeton Working Paper #393.

Conniffe, D., V. Gash and P.O'Connell (2000), "Evaluating State Programs: "Natural Experiments" and Propensity Scores," mimeo ESRI.

Corson, W. and W. Nicholson (1985) *Evaluation of the Charleston Claimant Placement and Work Test Demonstration*, Unemployment Insurance Occasional Paper 89-3, U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service.

Corson, W, P. Decker, S. Dunstan and A. Gordon (1989) *The New Jersey Unemployment Insurance Reemployment Demonstration Project: Final Evaluation Report*, Unemployment Insurance Occasional Paper 85-2, U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service.

Davidson, C and S. Woodbury (1993) "The Displacement Effect of Reemployment Bonus Programs," *Journal of Labor Economics*, vol. 11, no. 4 pp. 575-605.

Decker, P (1994) "The Impact of Reemployment Bonuses on Insured Unemployment in the New Jersey and Illinois Reemployment Bonus Experiment," *The Journal of Human Resources*, Vol. XXIX, no. 3

Dehejia, R.H and S. Wahba (1999), "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programmes," *Journal of the American Statistical Association* , Vol. 94, pp. 1053-1062.

Deloitte/Touche and A.Murphy (1998) Review of Community Employment Programme, Department of Enterprise, Trade and Employment.

Disney, R., A. Carruth, W. Franz, R. Jackman, R. Layard, H. Lehmann and J. Philpott (1992), Helping the Unemployed: Active Labour Market Policies in Britain and Germany. Anglo-German Foundation.

Dolado, J, F. Kramarz, S. Machin, A. Manning, D. Margolis and C. Tuelings (1996), "The impact of Minimum Wages in Europe," *Economic Policy*, pp. 319-372.

Dolton, P and D. O'Neill (1995), "The Impact of Restart on Reservation Wages and Long-Term Unemployment," *Oxford Bulletin of Economics and Statistics*, lvii, no. 4, pp. 451-471.

- Dolton, P and D. O'Neill (1996a), "Unemployment Duration and the Restart Effect: Some Experimental Evidence," *The Economic Journal* cvii, no. 435 (March), pp. 387-400.
- Dolton, P and D. O'Neill (1996b), "The Restart Effect and the Return to Full-Time Stable Employment," *Journal of the Royal Statistical Society, Series A:clix*, part 2, pp. 275-288
- Dolton, P. and D. O'Neill (1999) "The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Some Experimental Evidence from the U.K", mimeo, NUI Maynooth.
- Eissa, N and J. Liebman (1996), "Labour Supply Response to the Earned Income Tax Credit," *Quarterly Journal of Economics*, CXI, pp. 605-637.
- ESRI (1999) "Living in Ireland Research Group – Newsletter,"
- Feehan, S (1999), "Standard Hours Legislation and its effect on Employment: An Analysis of Canadian Data," MA Thesis, Economics Dept., NUI Maynooth.
- Franke, R and J. Kaul (1978) "The Hawthorne Experiments: First Statistical Interpretation," *American Sociological Review*, Vol. 43, no. 5, pp. 623-643.
- Freeman, R. (1998) "Worksharing to Full-Employment: Serious Option or Populist Fallacy," in R.Freeman and P.Gottschalk (eds) Generating Jobs: How to Increase the Demand for Less Skilled Workers, New York, Russell-Sage Foundation.
- Freeman, R and B. Hall (1986), "Permanent Homelessness in America," NBER Working paper no. 2013.
- Freeman, R (1990), "Employment and the Earnings of Disadvantaged Young Men in a Labor Shortage Economy," NBER Working Paper no. 3444.
- Friedlander, D, D. Greenberg and P. Robins (1997), "Evaluating Government Training Programs for the Economically Disadvantaged," *Journal of Economic Literature*, vol. XXXV, no. 4, pp. 1809-1855.
- Gorter, C and G. Kalb (1996) "Estimating the Effect of Counseling and Monitoring the Unemployed Using a Job Search Model," *Journal of Human Resources*, vol. 31, pp. 590-610.
- Greene, F, S. Machin and A. Manning (1996), "The Employer-Size Wage Effect: Can Dynamic Monopsony Provide an Explanation?," *Oxford Economic Papers*, Vol. 48, pp. 433-455.
- Hamermesh, D. and S. Trejo (2000), "The Demand for Hours of Labor: Direct Evidence from California," *Review of Economics and Statistics*, vol. LXXXII, no. 1, pp. 38-49.
- Hamermesh, D (1990), "Data Difficulties in Labor Economics," in E.Berndt and J. Triplett (eds) 50 years of Economic Measurement, University of Chicago Press.

Hamermesh, D (1999), "LEEping into the future of labor economics: the research potential of linking employer and employee data," *Labour Economics*, Vol. 6, no. 1, pp. 25-42.

Ham, J and R. LaLonde (1996) "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, vol. 64, no. 1, pp. 175-205.

Harmon, C. and I. Walker (1995), "Estimates of the Economic Return to Schooling for the United Kingdom," *American Economic Review*, Vol. 85, no. 5, pp. 1278-1286.

Heckman, J, and J. Smith (1995) "Assessing the Case for Social Experiments," *Journal of Economic Literature*, vol. 9 , no. 2 pp. 85-110.

Heckman, J., J. Smith and C. Taber (1998), "Accounting for dropouts in evaluations of social programs," *Review of Economics and Statistics*, vol. 80, no. 1, pp. 1-14.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, Vol. 46, no. 4, pp. 931-60.

Heckman, J. and R. Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (eds) Longitudinal Analysis of Labor Market Data, New York, Wiley.

Heckman, J (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47, no. 1, pp. 153-161.

Heckman, J., J. Smith and N. Clements (1997), "Making the Most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts," *Review of Economic Studies*, vol. 64, no. 4, pp. 487-535.

Heckman, J (1996) "Comment on : Identification of Causal Effects Using Instrumental Variables" *Journal of the American Statistical Association* , Vol. 91, pp. 459-462.

Heckman, J, H. Ichimura, J.Smith and P.Todd (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, Vol. 66, no. 5, pp. 1017-1098.

Heckman, J, R. LaLonde and J. Smith (1999) "The Economics and Econometrics of Active Labor Market Programs," forthcoming in Handbook of Labor Economics, Vol. III, eds O.Ashenfelter and D.Card.

Heckman, J. (1997), "Instrumental Variables," *The Journal of Human Resources*, Vol. 32, no. 3, pp. 441-462.

Hunt, J (1999), "Has Work-Sharing Worked in Germany ?," *Quarterly Journal of Economics*, Vol. CXIV, pp. 117-148.

Jenkins, J (1998), "Measurement of the Income Distribution: An Academic User's View," mimeo, University of Essex.

Johannesson, J and I. Persson-Tanimura (1978), "Labour Market Policy in Transition – Studies about the Effects of Labour Market Policy," EFA: Swedish Ministry of Labour, Stockholm.

Johnson, T.R and D.H Klepinger (1994) "Experimental Evidence on Unemployment Insurance Work-Search Policies," *Journal of Human Resources*, vol. 29, no. 3, pp. 695-717.

Katz, L (1996), "Wage Subsidies for the Disadvantaged," NBER Working paper no. 5679.

LaLonde, R (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, vol. 76, no. 4, pp. 604-620.

Lane, P. (1998), "Profits and Wages in Ireland, 1987-1996," *Journal of the Statistical and Social Inquiry Society of Ireland*, Vol. XXVII, Part V, pp. 223-247.

Leonard, J (1999), "Bringing the firm back in.," *Labour Economics*, Vol. 6, no. 1, pp. 43-52.

McCarthy, T (1998), "Tax and Progressivity in Ireland," paper presented to the Foundation for Fiscal Studies, December.

Maddala, G. (1983), Limited Dependent and Qualitative Variables in Econometrics, New York: Cambridge University Press.

Madden, D. (1999), "Relative or Absolute Poverty Lines: A New Approach," University College Dublin Working paper WP99/9.

Manser, M (1997), "Existing Labor Market Data: Current and Potential Research Uses," in J. Haltiwanger, M. Manser and R. Topel (eds) Labor Statistics Measurement Issues University of Chicago Press.

Meyer, B (1995a) "Lessons from the U.S Unemployment Insurance Experiments," *Journal of Economic Literature*, vol. XXXIII, pp. 91-131.

Meyer, B (1995b), "Natural and Quasi-experiments in Economics," *Journal of Business and Economic Statistics*, Vol. 13, no. 2, pp. 151-61.

Murphy, A. and B. Walsh (1996), "The Incidence of Male Non-Employment in Ireland," *Economic and Social Review*, Vol. 25, No. 5, pp. 467-490.

Murphy, A. and B. Walsh (1997), "Modeling the Duration of Male Unemployment," mimeo University College Dublin.

National Statistics Board (1998), Implementation of Strategy for Statistics, 1998-2002, Stationary Office, Dublin.

- Neumark, D. and W. Wascher (1997), "The New Jersey-Pennsylvania Minimum Wage Experiment: A Re-Evaluation Using Payroll Records," mimeo, Michigan State University.
- Nolan, B, D.O'Neill and J. Williams (1999) "The impact of A Minimum Wage in Ireland : Learning from a New Survey of Irish Firms," in The Impact of Minimum Wage in Ireland, Published by Department of Enterprise, Trade and Employment.
- Nolan, B (1998), "Low Pay in Ireland," in Report of the National Minimum Wage Commission, Vol. 2, Published by Department of Enterprise, Trade and Employment.
- O'Connell, P and F. McGinnity (1997), Working Schemes ? Active Labour Market Policy in Ireland, Ashgate Publishing Ltd., Aldershot.
- O'Connell, P (1999) "Assessing the Labour Market Impact of State-Sponsored Human Resource Interventions," Report to the Department of Finance, Dublin.
- O'Neill, D. and O. Sweetman, (1999) "Poverty and Inequality in Ireland 1987-1994 : A Comparison using measures of Income and Consumption" NUI Maynooth Working Paper N86/03/99
- Robinson, P (1995) "The Decline of the Swedish Model and the Limits to Active Labour Market Policy," CEPR Discussion paper no. 259.
- Swamy, P., R. Conway, and M. Le Blanc (1988a), "The Stochastic Coefficients Approach to Econometric Modeling: Part I: A Critique of the Fixed Coefficients Models," *The Journal of Agricultural Economics Research*, Vol. 40, no. 2, pp. 2-10.
- Swamy, P., R. Conway, and M. Le Blanc (1988b), "The Stochastic Coefficients Approach to Econometric Modeling: Part II: Description and Motivation," *The Journal of Agricultural Economics Research*, Vol. 40, no. 3, pp. 21-30.
- Swamy, P., R. Conway, and M. Le Blanc (1989), "The Stochastic Coefficients Approach to Econometric Modeling: Part III: Estimation, Stability Testing and Prediction," *The Journal of Agricultural Economics Research*, Vol. 41, no. 1, pp. 4-20.
- Van de gaer, D, N. Funnell and T.McCarthy, (1999), ""Statistical inference for two measures of inequality when incomes are correlated", *Economics Letters*, Vol. 64, pp.295-300.
- Vella, F. (1999), "Estimating Models with Sample Selection Bias," *Journal of Human Resource*, Vol. 33, no. 1 pp. 127-169.
- Wagner, G. (1998), "The German Experience A: an Economists viewpoint of prospects and some theoretical considerations for a better cooperation of academic and official statistics," mimeo Deutsches Institut fur Wirtschaftsforschung, Berlin.
- Walsh, P, E. Strobl and F. Barry (1998), "Aggregate Job Creation, Job Destruction and Job Turnover in the Irish Manufacturing Sector," *Economic and Social Review*, Vol. 29, no. 1, pp. 55-71.

White, M and J. Lakey (1992), "The Restart Effect: Does Active Labour Market Policy Reduce Unemployment," Policy studies Institute, London.

Williams, J and G. Hughes (1999), National Survey of Vacancies in the Private Non-Agricultural Sector 1998, published by the ESRI in association with FAS and Forfas.

Woodbury, S. and R. Spiegelman (1987) "Bonuses to workers and Employers to reduce Unemployment: Randomized Trials in Illinois," *American Economic Review*, vol.77, no. 4, pp.513-530.

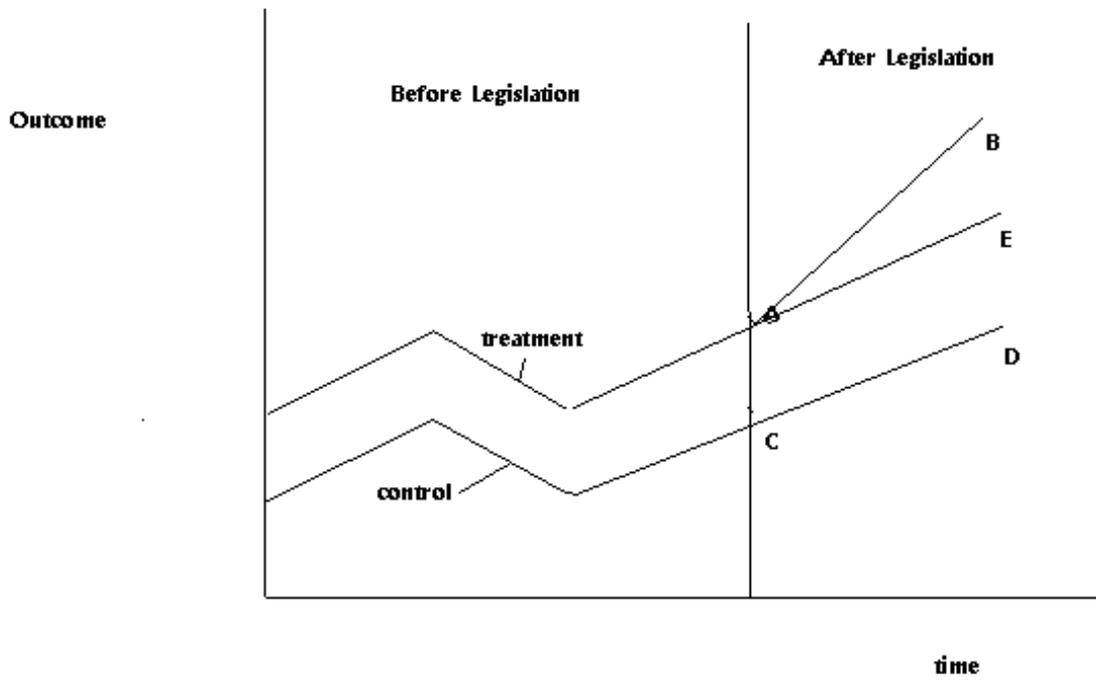


Figure 1: Alternative Program Evaluation Estimators

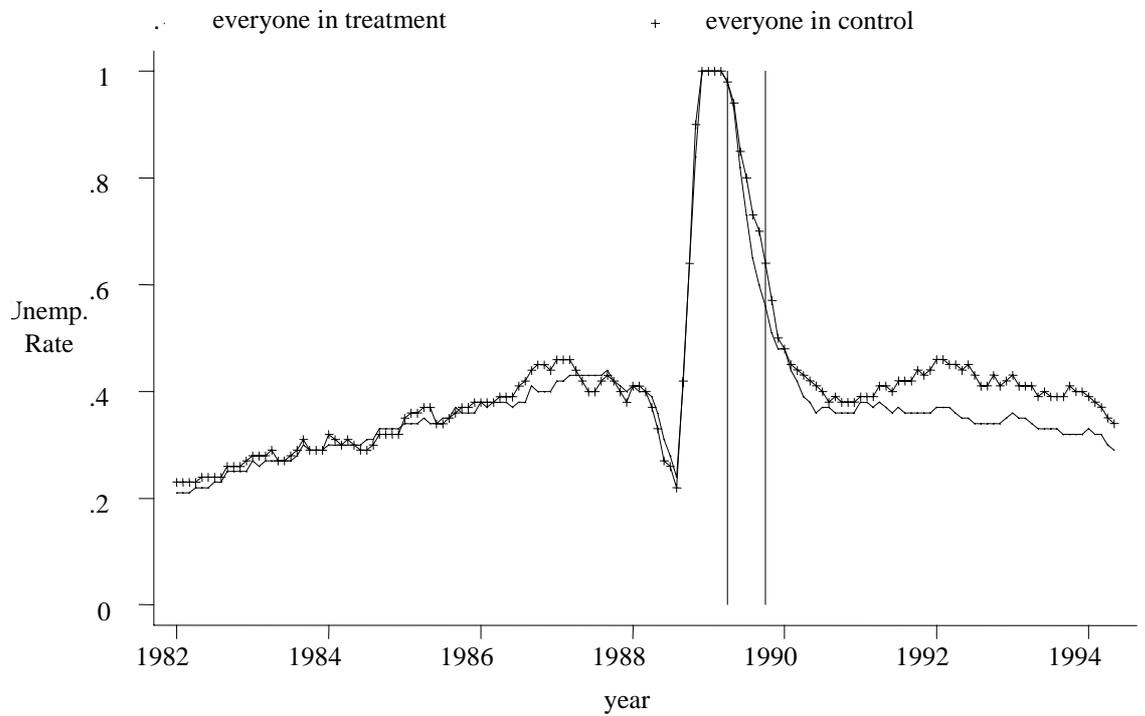


Figure 2: Unemployment Rates for the control and treatment groups in the Restart Experiment

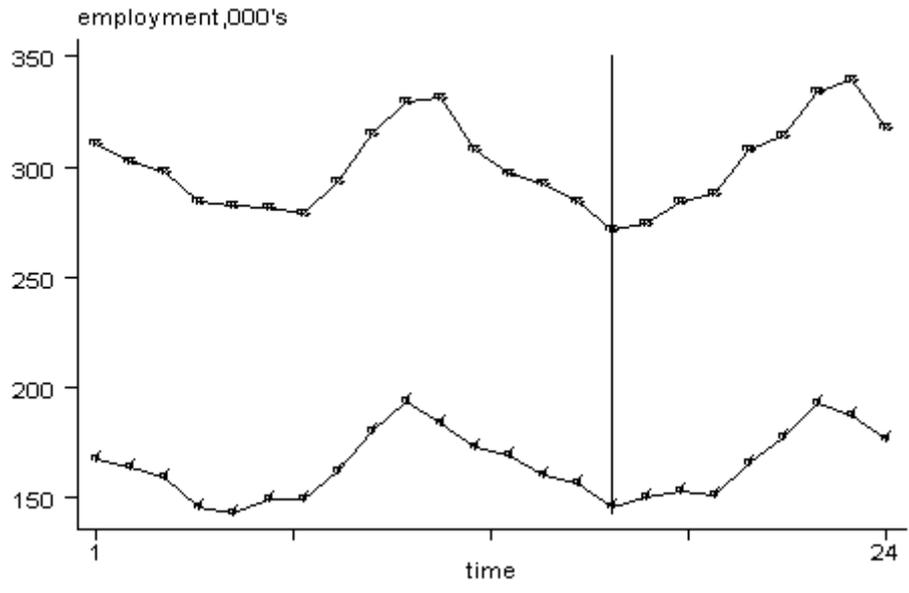


Figure 3: Full-Time Employment Levels for Newfoundland (nf) and Nova Scotia (ns)

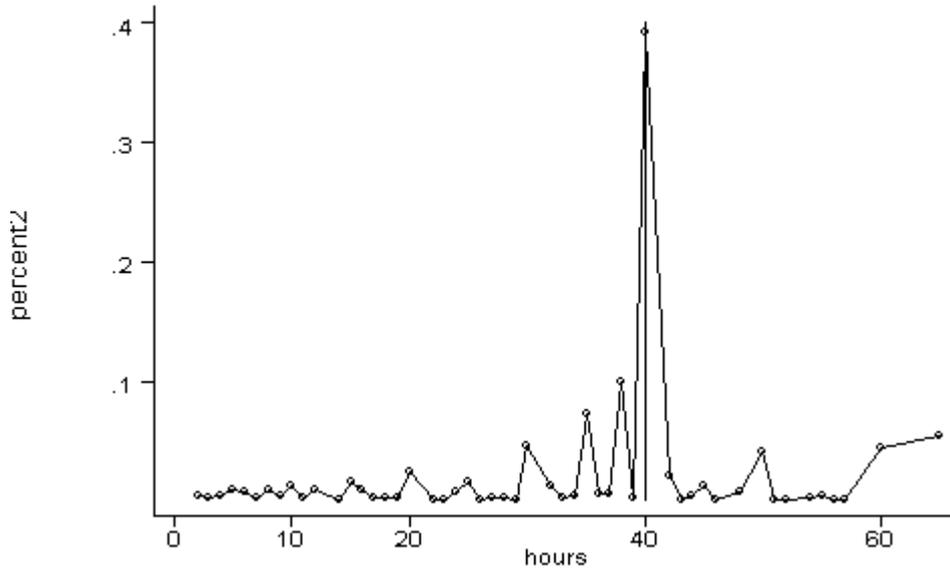


Figure 4a: Distribution of hours worked in Newfoundland in 1994

\*

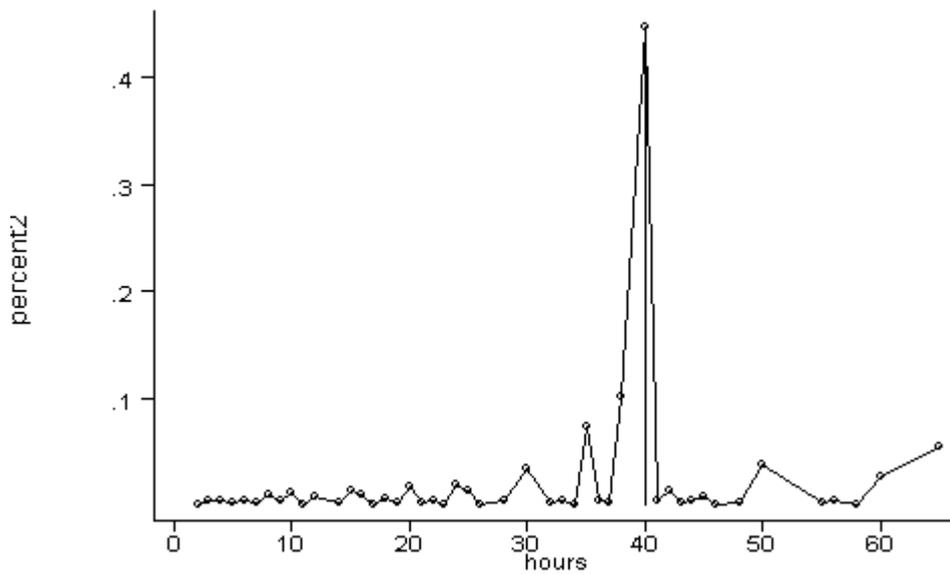


Figure 4b: Distribution of hours worked in Newfoundland in 1991

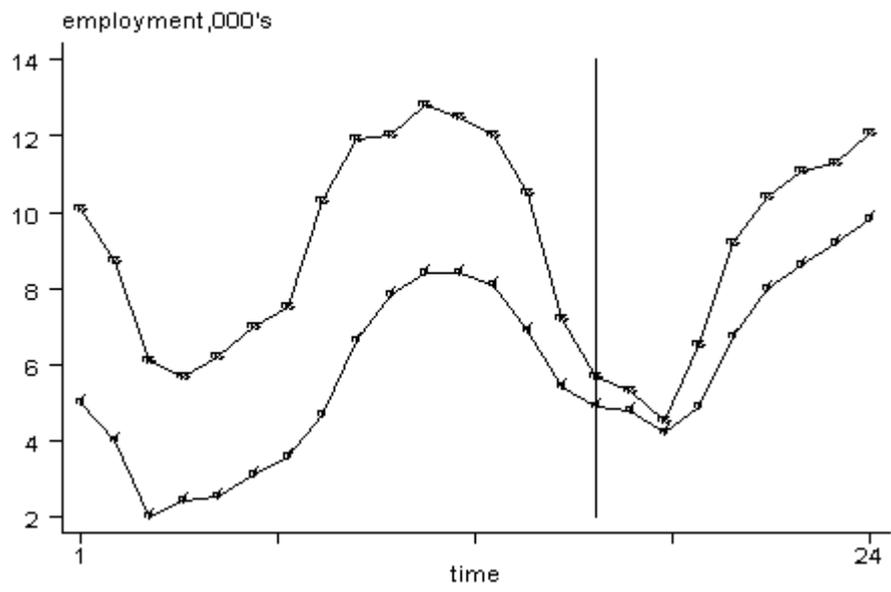


Figure 5: Employment Levels in Construction Industry for Newfoundland (nf) and Nova Scotia (ns)