

The Use of Evidence in Policy Making

Helsinki November 13th 2008

Richard Blundell

<http://www.ucl.ac.uk/~uctp39a/>

University College London and Institute for
Fiscal Studies

1

See also: The Mirrlees Review

Tax by Design: Reforming the Tax System for the
21st Century

Papers and commentaries available at
<http://www.ifs.org.uk/mirrleesreview>

2

The Use of Evidence in Policy Making

- Main question:
 - How should we best use evidence to *evaluate* policy reforms and to *improve* policy design?
- In this talk I am going to use three types of policies and three key alternative approaches to provide an answer.
- Drawing throughout from experience at the *Institute for Fiscal Studies* on the use of evidence in policy evaluation and in policy design - <http://www.ifs.org.uk/>

3

Three Approaches to Policy Evaluation

- Randomised-control experiments
 - Quasi-experimental evaluations
 - Micro(econometric)-simulations
 - Often thought of as competitors but each have their advantages and can be used to compliment each other in good evidence-based policy design.
 - Each exploit different types of data.
 - Each rely on different types of assumptions.
- See “Evaluation Methods for Non-Experimental Data”
<http://www.ifs.org.uk/fs/articles/0031a.pdf>

4

The Use of Evidence in Policy Making

- Three types of policy reform (all directed to the labour market attachment and earnings of the low skilled):
 - I. ERA: a bonus scheme for enhancing earnings and job retention for low paid single parents in work – *evaluated through a randomised control trial*
 - II. NDLP: a mandatory job search scheme for the young unemployed – *evaluated through an pilot-area based quasi-experiment*
 - III. WFTC: an earned income tax credit scheme for low income families – *evaluated (and improved) through a microeconomic simulation*
- All UK policies and in each case the evidence was used directly in the policy design

5

What Are we Trying to Achieve?

- Three slightly different objectives:
 1. A clear and accurate impact measure of a specific policy.
 2. An ability to examine alternative policy proposals.
 3. A mechanism for improving the design of policy – ‘optimal’ policy design.
- Never underestimate the importance of basic facts about the policy impact from careful microsimulation of first-round impacts.

6

The Use of Evidence in Policy Making

- What are we trying to achieve?
 - Randomised trials and quasi-experiments offer a clear and accurate average impact measure of a specific policy.
 - *Provided they are well designed.*
 - Micro(econometric)-simulation models offer a mechanism for evaluating alternative reform parameters and for improving policy design.
 - *But need to be convincing and robustly estimated.*
- First consider the ERA experimental evaluation

7

I. The ERA Policy Design and Evaluation

- Randomised-control trial and the ERA bonus experiment
- Ex-ante policy evaluation to answer the question:
 - Can a time-limited financial bonus given to low income workers, enhance their overall earnings and their labour market attachment?
- A specific policy proposal is chosen
- A specific target group is chosen
- A sample is drawn and a control group is derived by randomisation of those offered the ERA

8

I. The ERA Policy Design and Evaluation

- Policy: 2-year post-employment
- Employment retention bonus
 - £400, 3 times a year for 2 years
 - if in full-time work (≥ 30 h pw) for ≥ 13 out of every 17 weeks
 - If qualify for the earned income tax credit – WTC
- Training bonus
 - tuition up to £1,000
 - training bonus of £8 per hour (max £1,000) upon successful completion
 - if works ≥ 16 h/w

9

The Experimental Design

- large-scale random assignment: - 16,000 individuals
- Key information on all study participants at intake:
 - Demog, socio-economic, education
 - Work history
 - Current work status (hours and earnings if working)
 - Driving licence and access to car
 - Childcare, health, housing, basic skills, transport
- Administrative Data (WPLS)
 - History: employment, benefit
 - Outcomes: employment, benefit, earnings

10

Balance Across Treatment and Control

	Treatment	Control
Age		
Under 30	15.5	15.8
30-39	49.2	49.6
40 or older	35.3	34.5
Education Qualifications		
None	14.1	13.7
GCSE	49.9	45.1*
A-level and College	36.1	41.2*
Number of Months Worked		
1-12	14.8	14.4
13+	85.2	84.6

11

The Estimated Impacts of ERA

- The estimated impact show a significant impact on earnings and a significant impact on full time work

	Treatment	Control	Difference	%
Average Total Earnings	17267	16392	874**	5.3
Ever Worked Full-time	41.5	30.0	11.5***	38.2
Worked Full-time 4 consecutive months	38.2	28.9	10.3**	35.8

- Major impact on moving toward time-limited bonuses for earnings retention and enhancement

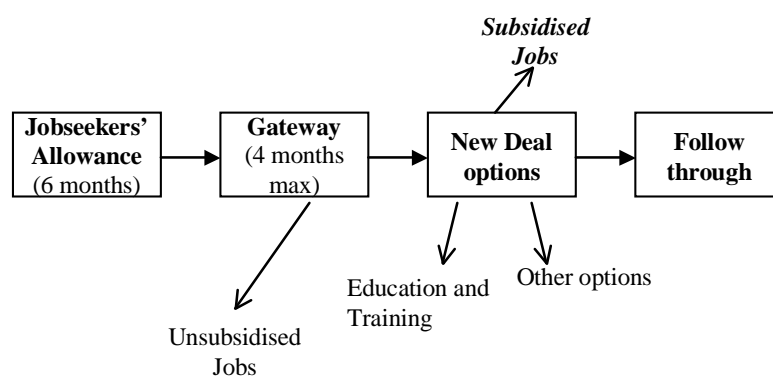
12

II. The New Deal for the Young Unemployed- UK

- A *conditional and time-limited* welfare-to-work program.
- eligibility depends on being aged 18-24 – mainly singles
- eligibility conditional on 6 months unemployment
- mandatory with job search monitoring for four months
- those not moving into unsubsidised employment move in to options:
 - Wage subsidy provided to the employer which is also time limited – 6 months
 - Other options

13

II. The New Deal for the Young Unemployed- UK



See “Evaluating the Employment Impact of a Mandatory Job Search Assistance Program,” <http://www.ucl.ac.uk/~uctp39a/Blundell-JEEA.pdf>

14

II. The New Deal for the Young Unemployed

- No experimental research design but an *area based treatment and control*
 - use of Control/Pilot Areas
 - compare these areas
 - ‘perfect’ quasi-experiment
 - use differences-in-differences with matching
- *Mandatory* – so that all are treated
- *24 age cut-off* provides a further ‘natural’ comparison’ group
- Can assess *spill-over and GE effects*

The New Deal for the Young Unemployed

- *Comparison groups (before and after reform):*
- Compare 19-24 year olds in pilot areas and control areas
 - *Pure diff-in-diff effect*
- Compare 19-24 year olds in pilot with older groups in pilot areas
 - *Spill-overs to the older group?*
- Compare all groups in pilot areas and matched control areas
 - *Overall ‘GE’ (general equilibrium) effects*

16

New Deal Impacts - Outflow to Employment

		Estimates based on the Difference in Differences with Matching	
Treatment group	Comparison group	Linear Matching (OLS/Linear probability model)	Propensity score matching using smoothing splines
19-24 year olds living in Pilot areas	19-24 year olds living in all non-Pilot areas	0.110** (0.039)	0.104** (0.046) (0.024;0.182)
19-24 year olds living in Pilot areas	25-30 year olds living in Pilot areas	0.104* (0.055)	0.078 (0.079) (-0.050;0.195)

Source: Blundell et al, JEEA, 2005 ¹⁷

The New Deal for the Young Unemployed

- A well design are pilot and control can produce reliable evaluation reports
- If policy is mandatory and a well matched control group can be found in the comparison areas then can recover a useful estimate of the average impact on those who received the policy.
- Comparing all (relevant) groups in pilot areas and matched control areas can also be used to assess the importance of GE and spill-over effects.

18

III. Tax Credit reform in the UK

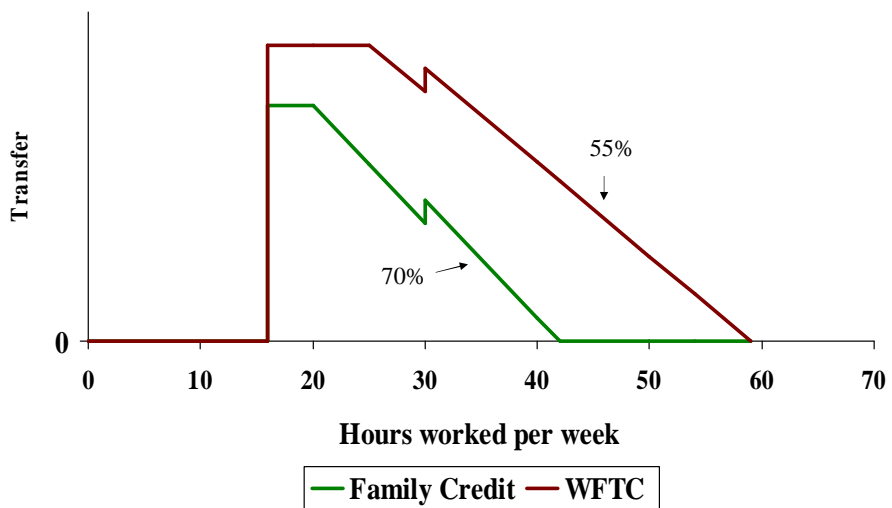
- FC (family credit) before 2000, expanded early in 1990s
- WFTC (working families tax credit) reform in 2000, and subsequent expansions in 2002
 - influenced by the success of the EITC expansion in the US
 - especially generous to families with young children
- Micro-simulation evaluation

See “The Labour Market Impact of the Working Families’ Tax Credit”,
<http://www.ucl.ac.uk/~uctp39a/blundell-WFTC-Fiscal%20Studies.pdf>

Three eligibility criteria for WFTC:

- work eligibility
 - e.g. 16 or more hours per week in the UK
- family eligibility
 - e.g. children in full time education
- income eligibility
 - e.g. family net income below a certain threshold then eligible for adult credit plus age-dependent amounts for each child
 - if income is above the threshold then the amount of credit is tapered away, e.g. 55% per extra pound of net income
 - joint rather than individualised taxation

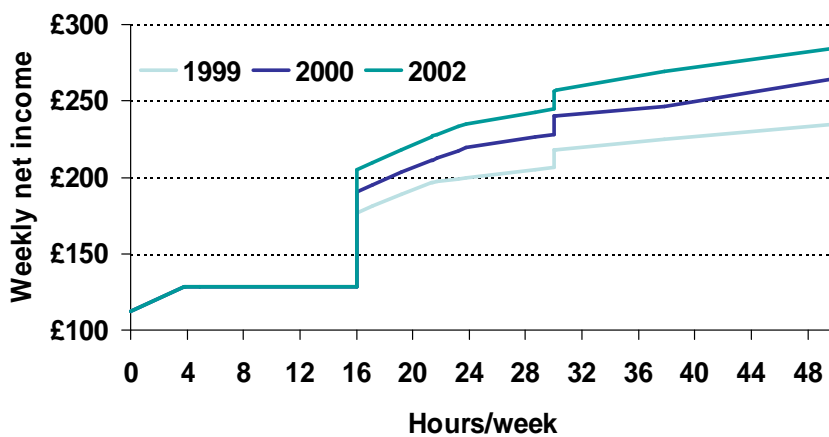
The WFTC Reform



transfers per week for a min. wage lone parent

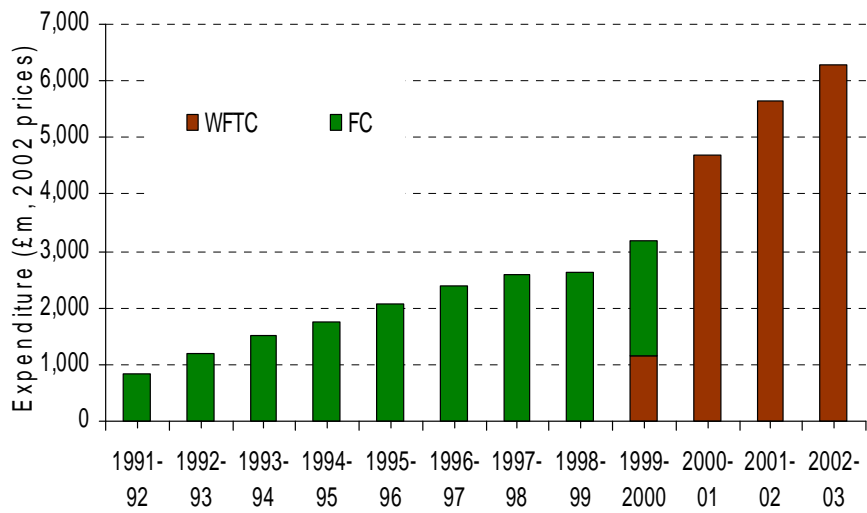
Impact of WFTC on lone parent, 2 children

Changes in the Budget Constraint



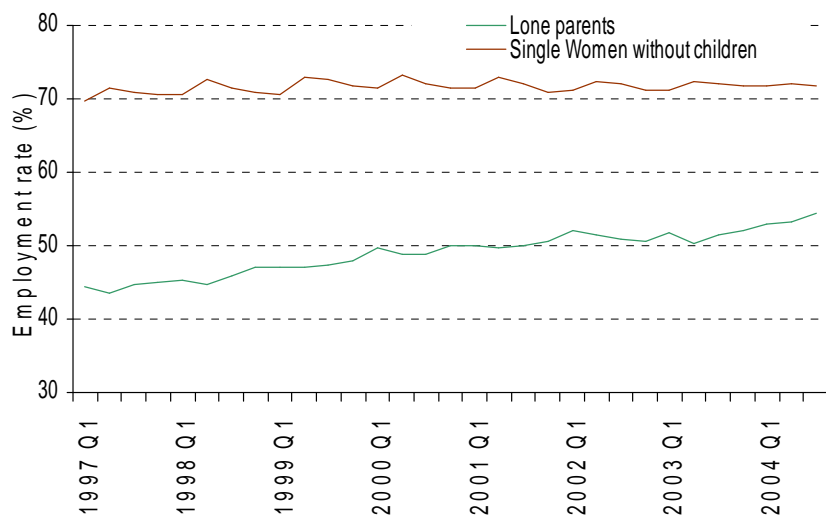
22

Expenditure on in-work credit programmes in the UK



23

Can the reforms explain the change in employment rates?



24

WFTC Evaluation: Matched Difference-in-Differences

Employment Rate

<i>Single Women</i>	Marginal Effect	Standard Error	Sample Size
Family Resources Survey	3.5	1.55	25,163
Labour Force Survey	3.6	0.55	233,208

Data: FRS, 45,000 adults per year, Spring 1996 – Spring 2002.

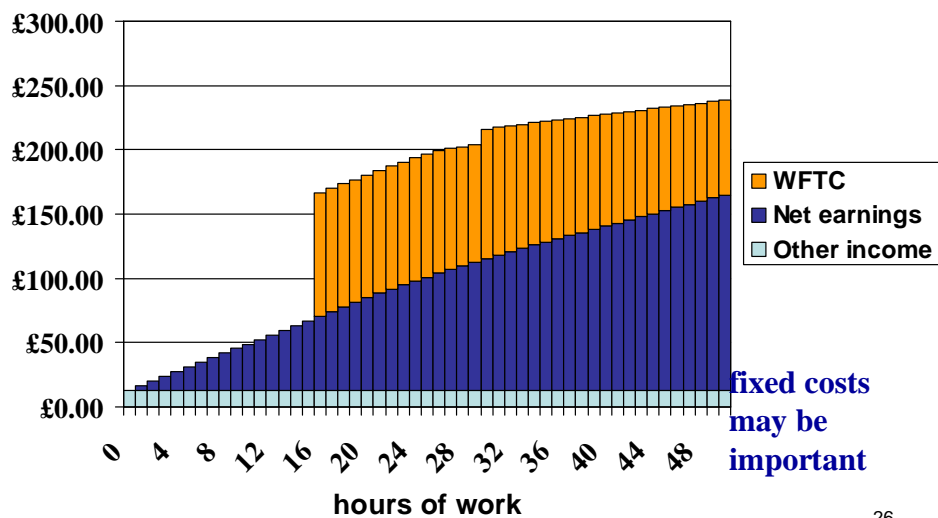
Outcome: employment. Average impact x 100, employment percentage.

Matching Covariates: age, education, region, ethnicity,..

Drop: Summer 1999 – Spring 2000 inclusive

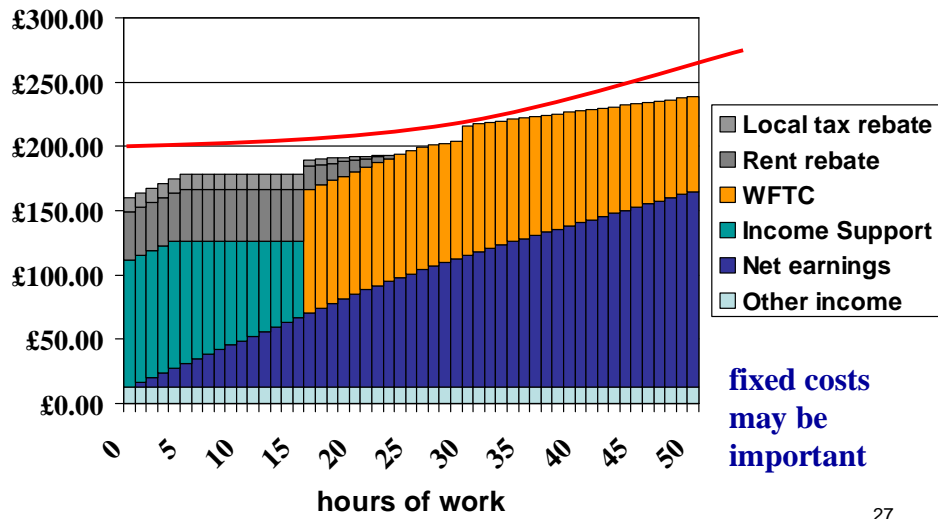
25

The interaction with other benefits



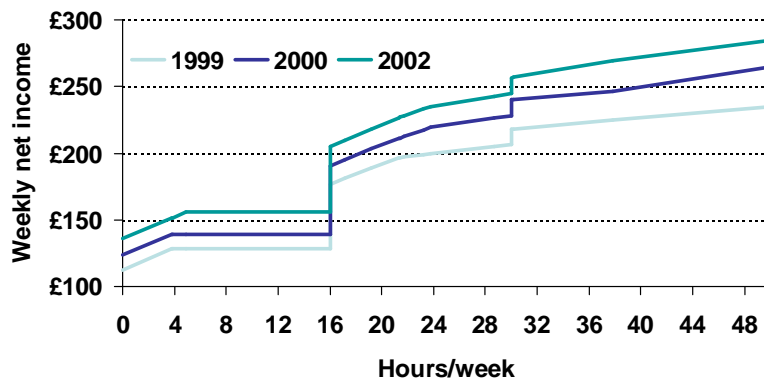
26

The interaction with other benefits



27

Impact of WFTC & increases in welfare benefit on lone parent, 2 children



28

Structural Micro-Simulation Results:

WFTC Expansion – ‘large’ impact

	All	y-child 0 to 2	y-child 3 to 4	y-child 5 to 10	y-child 11 to 18
Change in employment rate:	5.95	3.09	7.56	7.54	4.96
	<i>0.74</i>	<i>0.59</i>	<i>0.91</i>	<i>0.85</i>	<i>0.68</i>
Average change in hours:	1.79	0.71	2.09	2.35	1.65
	<i>0.2</i>	<i>0.14</i>	<i>0.23</i>	<i>0.34</i>	<i>0.2</i>

Notes: Simulated on FRS data; Standard errors in italics.

All: 5.12 without change in take-up – key impact effect

29

Structural Micro-Simulation Results:

All Reforms

	All	y-child 0 to 2	y-child 3 to 4	y-child 5 to 10	y-child 11 to 18
Change in employment rate:	3.66	0.65	4.53	4.83	4.03
	<i>0.84</i>	<i>0.6</i>	<i>0.99</i>	<i>0.94</i>	<i>0.71</i>
Average change in hours:	1.02	0.01	1.15	1.41	1.24
	<i>0.23</i>	<i>0.21</i>	<i>0.28</i>	<i>0.28</i>	<i>0.22</i>

- matches with the quasi-experimental results
- shows the structural model predictions are quite accurate

Notes: Simulated on FRS data; Standard errors in italics.

30

Interpreting the Employment Impact Results

- Small effects are due to interaction of WFTC with other taxes and benefits rather than ‘small’ response elasticities.
 - and the rise in family allowances (all reforms)
- Also shows the structural model predictions are quite accurate
- Can use the model to simulate alternative policy reforms
 - Changes to the parameters of the reform
 - Changes to the characteristics of the eligible population

31

Interpreting the Employment Impact Results

- Can use the model elasticities to evaluate the ‘optimal’ design
- An ‘optimal’ design is one where there are relatively lower marginal tax rates at places where labour supply elasticities are relatively high
- Use this insight to redesign the tax schedule for low wage workers
- Points to tax credits as being optimal and generally a lowering of the marginal tax rates and participation on lower wage workers.

32

Structural Model Elasticities

(a) Youngest Child Aged 11-18

<i>Earnings</i>	<i>Density</i>	<i>Extensive</i>	<i>Intensive</i>
0	0.3966		
80	0.1240	0.5029	0.5029
140	0.1453	0.7709	0.3944
220	0.1723	0.7137	0.2344
300	0.1618	0.4920	0.0829
<i>Participation elasticity</i>		1.1295	

33

Structural Model Elasticities

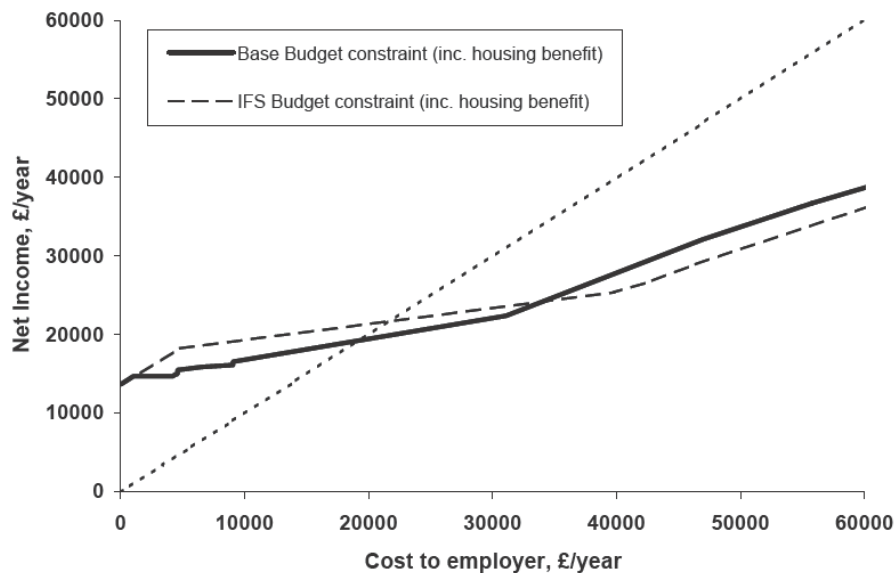
(c) Youngest Child Aged 0-4

<i>Earnings</i>	<i>Density</i>	<i>Extensive</i>	<i>Intensive</i>
0	0.5942		
80	0.1694	0.2615	0.2615
140	0.0984	0.6534	0.1570
220	0.0767	0.5865	0.1078
300	0.0613	0.4984	0.0834
<i>Participation elasticity</i>		0.6352	

- Implications for the optimal schedule

34

Tax Credit Reform, budget constraint, lone parents



The Use of Evidence in Policy Making

What have we learned?

- Convincing evidence-based research must prove itself robust – see Mirrlees Review for examples.
- Randomised trials and quasi-experiments offer a clear and accurate average impact measure of a specific policy.
- Limited in what they can measure and cannot be used for looking at sets of alternative policy proposals.
- Micro(econometric)-simulation models offer a mechanism for evaluating alternative reform parameters and for improving policy design.
- Matching tax reform to response elasticities.
- *Use a baseline (quasi-)experiment to check robustness*³⁶