

Estimating Models of Benefit Take-up

Mike Brewer*

Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE

December 15, 2003

Abstract

This paper summarises the economic literature on modelling take-up, with a focus on income-related or work-related benefits or tax credits. It gives a framework for thinking about aggregate measures of take-up rates, and shows how existing studies differ from each other in their data requirements and modelling complexity. It also shows how models of non-take-up, when appropriate data are available, can account for measurement and expectational errors, and how they can value the costs of claiming and receiving some benefits.

1 Introduction

Whether an individual claims a benefit or tax credit or participates in a government programme to which they are entitled seems like a straight-forward decision that should be simple to model. But many studies of government programmes assume full take-up - that everyone who is entitled claims or participates - and there is little economic research on what determines take-up (or the lack of it). To focus on a particular area, accounting for

*©Crown Copyright 2003. Published by Inland Revenue 2003. ISBN 1-904983-07-3. This work was part-funded by the Inland Revenue under the project "Econometric Research: Impact of Working Families' Tax Credit", and part-funded by the ESRC Research Centre for the Microeconomic Analysis of Public Policy. It is partly intended as background to currently on-going work on the impact of the WFTC on take-up and labour supply. I am grateful to Richard Blundell, Alan Duncan, Andrew Shephard, María José Suárez, Ian Walker and Yu Zhu for helpful discussions and/or collaboration on work which informed this. This paper was written at the same time as an ESRC-funded project was investigating take-up of the MIG amongst pensioners in the UK (see <http://www.le.ac.uk/economics/sep2/stigma.html> for a summary). All errors and omissions remain mine alone.

non-take-up is vital when estimating labour supply models, as the financial reward from work will be over-estimated if we assume that all in-work support is taken up¹; and studying the determinants of non-take-up is important as that might tell us whether someone is not working because they prefer not to work, or because they prefer not to claim in-work support.

In the light of these and other factors, the aim of this paper is to review how non-take-up has been studied by economists, focusing on examples from the labour supply literature. It examines whether there are limits on what one can learn about non-take-up from survey data given the presence of modelling, measurement and survey errors, and identifies methods in the literature that have been used to overcome these. It also shows how economic models can produce estimates of the utility cost of claiming benefits.² In the course of doing this, it will be possible to suggest what data one would ideally need if one wanted to learn precisely why people did not apply for benefits to which they were entitled.

The paper is organised as follows. By way of further motivation and introduction, Section 2 discusses what is non-take-up, and presents an overview of the literature. Section 3 looks at estimates of the aggregate take-up rate, and the biases that could emerge when these are based on survey data. Section 4 presents a framework for economists' analysis of non-take-up, and analyses the informational requirements of various approaches in the literature. Particular studies of non-take-up are analysed in section 5. Section 6 concludes. This paper was written at the same time as an ESRC-funded project was investigating take-up of the MIG amongst pensioners in the UK (see <http://www.le.ac.uk/economics/sep2/stigma.html> for a summary).

¹It is not easy to say what sort of bias would result, though, as it will depend on what the form of the means-tested benefit. But, for example, if full take-up of an in-work was assumed, then people observed not working would seem to be less responsive to financial incentives than they actually were, and this might increase the estimate of the fixed costs of work.

²This paper is concerned with take-up of cash benefits, but most of the concepts apply to measuring take-up of any voluntary government program. From now on, we use the terms "take-up" (or "claim") and "benefit", and will let the reader substitute "participate" and "programme" (or "tax credit") at will.

2 Why do take-up studies ?

Non-take-up simply describes a situation where someone does not claim a benefit to which they are entitled. Empirical studies of non-take-up typically compare data on receipt of benefits recorded in household surveys (or as recorded in matched administrative data sources) with data on entitlement for benefits produced by a microsimulation model operating on data on household characteristics from the same household survey. As we elaborate below, non-take-up is rationalised by assuming that there are some costs to claiming or receiving the benefit. Past economic studies of non-take-up have looked at several related areas of interest:³

1. The largest section of the literature has estimated the probability that entitled families actually claim a benefit using household-level data. These can be grossed up to derive the overall take-up rate, or the proportion of the total potential expenditure that is being claimed. This is usually thought of as being the “headline” summary measure of take-up. Examples of the literature include Fry and Stark (1993), DWP (2003) and Hancock et al (2003). An obvious variant is the proportion of expenditure that is claimed, which is the same measure weighted by entitlement (see Wicks (2003)).

There are a number of reasons why this statistic is of interest. It is an intuitive indicator of how well a benefit is reaching its intended population, assuming that the intended population is “everyone who is entitled to it”.⁴ It is easy to calculate, to understand and to compare with other benefits over time or across countries. Accurate forecasts of caseload and expenditure take-up improve the accuracy of public spending forecasts. Finally, the proportion of truly-entitled families who take-up a benefit should be one of the success measures of a means-tested benefit, because there is usually the alternative of a lump-sum or otherwise automatic transfer payment. The main political justification in the UK for using means-tested benefits is that they reduce the cost to government of achieving a given amount of redistribution which

³A recent survey article is Remler et al (2001).

⁴Yaniv (1997) and Besley and Coates (1992) offer different suggestions why it may be optimal to have less than full take up.

could also be achieved through non-means-tested benefits.⁵ For that reason, take-up of means-tested benefits ought to be compared to the (typically very high) take-up rates of non-means-tested benefits. One drawback of the overall take-up rate is that, although it is a good measure of the success of initiatives that try to increase take-up, it is not always informative when the size of the entitled population has changed, perhaps through a change in the rules determining entitlement.

2. A natural follow-up is to ask what factors help explain or predict (in a regression framework) whether an entitled family would take-up a benefit. Examples of this literature include Blundell et al (1988), Fry and Stark (1993), Duclos (1995 & 1997), Riphahn (2001), Hancock and Barber (2003), Brewer et al (2003a). Few of these studies claim to be estimating structural models of non-take-up (in other words, they do not estimate a utility function). Instead, by indicating which observable characteristics are linked to non-take-up, they may add to our understanding of what might cause non-take-up.⁶ This approach has advantages over the first when examining how take-up changes when the a benefit increases in generosity, such as when WFTC replaces FC, because it controls for other characteristics.
3. Some labour supply studies have sought to model take-up of benefits simultaneously with labour supply behaviour. Usually, the focus has been on improving the accuracy of the labour supply equation, rather than on learning about non-take-up and stigma costs. Examples of this literature include: Moffitt (1983), Hoynes (1996), Bingley and Walker (1997), Keane and Moffitt (1998), Blundell et al (1999 & 2000), Andren (2003).

Two add-ons to the methods described above are:

⁵This is particularly put forward for families with children and pensioners, who who a universal (ie non-means-tested) benefit exists. It is less applicable when thinking about means-tested benefits for people of working age without children.

⁶As noted in Remler et al (2001) there may be problems interpreting some of the coefficients. For example, high education levels tends to be positively correlated with non-take-up. Education probably lowers the information cost of applying, but might proxy social class which could be related to pure stigma effects. Education is also correlated with assets and "permanent income", both of which are measured badly by surveys and probably important determinants of preferences for income and eligibility for benefits .

- to allow for modelling errors (on the part of the analyst) or expectational errors (on the part of the claimant) when estimating take-up
- to estimate a value (in some metric) of the stigma costs of receiving benefits.

Both add complexity. Some studies in the first two categories have tried to deal with modelling errors, and, depending on the precise specification, it is possible for studies in the last 2 categories to value the stigma cost.

There are also qualitative studies that investigate the stated reasons behind non-take-up of benefits, such as McConaghy et al (2003), but we do not review them here.

3 Measuring aggregate non-take-up

3.1 What is non-take up?

To consider this further, below we outline a common framework, drawing on Duclos (1995). There are three agents involved in a study of non-take-up: the individual, the government agency that administers the benefit, and "the analyst" (a composite figure who both collects the survey data and runs the tax-benefit model). The government agency has an information set Ω^g (information from the benefit application form, and the agency's knowledge and interpretation of the benefit rules), and the analyst has an information set Ω^a (data from a household survey and the analyst's interpretation of the tax and benefit rules), where Ω^g and Ω^a are possibly overlapping subsets of Ω^* , the set of all information.

Define B^* to be the true entitlement of a family as determined by the means-tested benefit rules enshrined in legislation and regulations. Call the actual entitlement received B^g where:

$$B^g = E \{B^* | \Omega^g\}$$

In general, B^g differs from B^* because:

- the agency may measure household characteristics with error

- the agency may make errors when interpreting the benefit rules laid down in legislation (the fact that some appeals against benefit or tax credit decisions are successful indicates that this happens sometimes)

The analyst typically uses a microsimulation model and data from a household survey to produce an estimate of the entitlement B^a . This will differ from B^* because:

- measurement error or sampling defects may mean that surveys measure household characteristics with error
- there may be changes in circumstances between the time of the survey and the time when an individual's claim for benefits was last assessed.
- the analyst may make errors in interpreting the benefit rules laid down in legislation

So:

$$B^a = E \{B^* | \Omega^a\}$$

As $\Omega^g \neq \Omega^a$, the analyst's estimate is not, in general, the same as the agency's.⁷

If $R = 1$ denotes receipt of a benefit by a family (we assume that R is in Ω^g and Ω^a), then we can partition the population into four sets of interest:

eligible recipients (ER)	$\{R = 1, B^* > 0\}$
eligible non-recipients (ENR)	$\{R = 0, B^* > 0\}$
non-eligible recipients (NER)	$\{R = 1, B^* \leq 0\}$
non-eligible non-recipients (NENR)	$\{R = 0, B^* \leq 0\}$

The analyst will have to estimate these sets based on Ω^a :

⁷So B^g differs from B^a because:

- the agency and the survey may measure household characteristics differently
- the agency and the analyst may make different errors when interpreting the benefit rules laid down in legislation
- the analyst's information on household characteristics may not cover the same time period as the agency's.

eligible recipients (ER^a)	$\{R = 1, B^a > 0\}$
eligible non-recipients (ENR^a)	$\{R = 0, B^a > 0\}$
non-eligible recipients (NER^a)	$\{R = 1, B^a \leq 0\}$
non-eligible non-recipients ($NENR^a$)	$\{R = 0, B^a \leq 0\}$

3.2 The bias in estimating aggregate take-up rates

A good measure of take-up, T^* , would be the probability (or expectation) that a truly-entitled family receives the benefit, ie:⁸

$$T^* \equiv P(R = 1 | B^* > 0) = E(T^* | \Omega^*) = \frac{P(R = 1 \& B^* > 0)}{P(B^* > 0)}$$

Define T^a as the estimate of T^* made by the analyst:

$$T^a \equiv P(R = 1 | B^a > 0) = E(T^* | \Omega^a) = \frac{P(R = 1 \& B^a > 0)}{P(B^a > 0)}$$

How do these compare? In fact, we can say little in general terms. To see this, define p^g as the probability that a recipient family is not truly entitled (so the probability that the agency has wrongly awarded a benefit), p^a is the analyst's equivalent of this (the probability that a recipient unit is deemed not to be entitled by the analyst):

$$p^g \equiv \frac{P(R = 1 \& B^* = 0)}{P(R = 1)}$$

$$p^a \equiv \frac{P(R = 1 \& B^a = 0)}{P(R = 1)}$$

Some algebra shows that:

$$T^a = T^* + \left(\frac{P(B^* > 0)}{P(B^a > 0)} - 1 \right) T^* - (p^a - p^g) \frac{P(R = 1)}{P(B^a > 0)},$$

so we cannot even say in general whether T^a is an over- or under-estimate of the true take-up rate, T^* . If we assume that that the analyst correctly estimates the proportion of the population who are entitled, ie: $P(B^* > 0) = P(B^a > 0)$, then the expression above simplifies to:

⁸This draws on Duclos (1995).

$$T^a = T^* - (p^a - p^g) \frac{P(R=1)}{P(B^* > 0)} .$$

So, even if the analyst gets the proportion entitled to a benefit correct, there may still be a bias in the take-up rate estimate, and the greater the inaccuracy of the analyst's estimate of entitlement relative to the agency's, the more the analyst's estimate of take-up underestimates the true rate. p^g is unknown, and $P(B^* > 0)$ is only known by the assumption that it equals $P(B^a > 0)$, so we can never calculate the actual bias.⁹

Some analysts use another estimate of take-up that includes the non-entitled recipients in the denominator.¹⁰ By increasing the denominator and numerator by the same amount, this always leads to a higher estimate of take-up. There is no intuitive way of thinking about this measure, but define the "true" rate T_{iner}^* :

$$\begin{aligned} T_{iner}^* &\equiv \frac{P(R=1)}{P(R=0 \& B^* > 0) + P(R=1)} \\ &\stackrel{ie}{=} \frac{P(R=1)}{(1-T^*) \cdot P(B^* > 0) + P(R=1)} \\ &\stackrel{ie}{=} \frac{T^* \cdot P(B^* > 0) + N^* \cdot P(B^* = 0)}{P(B^* > 0) + N^* \cdot P(B^* = 0)} \end{aligned}$$

where N^* is the probability that someone who is not entitled is receiving the benefit, ie:

$$N^* \equiv P(R=1 | B^* = 0)$$

The analyst's estimate of this is:

$$\begin{aligned} T_{iner}^a &= \frac{P(R=1)}{P(R=0 \& B^a > 0) + P(R=1)} \\ &\stackrel{ie}{=} \frac{P(R=1)}{(1-T^a) \cdot P(B^a > 0) + P(R=1)} \\ &\stackrel{ie}{=} \frac{P(R=1)}{\left(1 - \left(T^* \left(\frac{P(B^* > 0)}{P(B^a > 0)}\right) - (p^a - p^g) \frac{P(R=1)}{P(B^a > 0)}\right)\right) \cdot P(B^a > 0) + P(R=1)} \end{aligned}$$

⁹In a meta-analysis of take-up studies, Remler et al (2001) suggest that estimates based on survey data tend to produce lower take-up rates than those based exclusively on administrative data, suggesting that $p^a > p^g$.

¹⁰See DWP (2003), for example.

As above, without any more information, we cannot sign the bias to this, but if $P(B^a > 0) = P(B^* > 0)$ then T_{iner}^a will also under-estimate the true rate of T_{iner}^* the greater the inaccuracy of the analyst's measure of entitlement relative to the agency.

3.3 Examples in the literature

Using formulae such as those shown above, DWP and its predecessors have produced estimates of the caseload and expenditure take-up of the key means-tested benefits in the UK for some years (see DWP (2003) or Inland Revenue (2002), for example). Their preferred estimate of aggregate take-up is T_{iner}^a , where they estimate the number of recipients - $P(R = 1)$ - from administrative sources (this should be unbiased), and estimate ENR^a from survey data (FRS or FES in practice). Their estimates of Family Credit take-up include an adjustment to the raw estimate of ENR^a to take account of the six-month award period of FC.¹¹

Fry and Stark (1993) examined take-up trends (both T^a and T_{iner}^a) of various means-tested benefits over the 1980s. They used the FES to estimate ER^a , NER^a and ENR^a . They chose to use the FES, rather than administrative data, to estimate the number of recipients so that they could then estimate a model of non-take-up consistent with their aggregate estimates, and so that they could distinguish between ER^a and ENR^a . Researchers using the PRILIF data-set have used that data to estimate aggregate take-up rates (eg Finlayson and Marsh (1998)). More recently, researchers have used data from FACS to estimate comparable take-up rates for FC and WFTC (see Marsh et al (2001) and McKay (2003), as well as Brewer et al (2003a)). These studies, like Fry and Stark (1993), used only survey data to estimate T^a .

¹¹This is because the set of FC recipients is the set of people who applied and were assessed to be eligible for FC at some point in the past 6 months. To be comparable with this, their estimate of ENR^a needs to estimate the number of non-recipients who would have been eligible had they made a claim for FC on any day in the past 6 months. They first estimate the number of non-recipients who would have been eligible based on the circumstances in the survey, and then scale this upwards using an adjustment factor based on the 6-monthly renewal rate for FC awards by family type (so this assumes that the income fluctuations of ENR are identical to those of recipients). Their estimates of take-up are clearly sensitive to this parameter. For example, the take-up rates for FC in 1996/7 assumed that for every identified ENR family, 1.23 couple families (1.18 lone parents) would have been an ENR on any day in the past 6 months. A 20% increase in this factor reduces the take-up estimate for couples from 67% to 62% and from 79% to 75% for lone parents.

4 Economic models of non-take-up

This section discusses some of the theoretical approaches that have been used to rationalise non-take-up, and then outlines the data requirements for estimating such models. The following chapter then describes how different studies have empirically estimated these models.

4.1 “Optimal non-take-up” with no modelling errors

The framework first outlined in Moffitt (1983) for analysing non-take-up suggests that people do not take-up benefits if the disutility of claiming and receiving the benefit outweigh the utility gain of the extra income. The programme considered in Moffitt (1983) - Aid to Families with Dependent Children - involved a form of job-search activity by claimants, which we can clearly identify as a cost of programme participation or take-up. In the UK, job-search requirements are less important, so the hypothesised disutility might be information costs (awareness of scheme, complexity of forms), process costs (time requirements), or outcome costs (“stigma”).¹² In a simple case, with exogenous income y , and assuming utility is separable in stigma costs¹³, families will take up some benefit which is worth $B^*(y; X)$ if:

$$U(y + B^*(y; X); X) - C(y; X) > U(y; X)$$

where y is original income, X describes the family characteristics, $C(y, X)$ is the utility cost of claiming and receiving the benefit, and we assume that families costlessly know their entitlement, and that there are no errors (ie $B^* = B^g$). This leads to the following observation rule:

where the objects of interest are $U(y, X)$ and $C(y; X)$. If there is more than one benefit than can potentially be taken-up, then the utility function can be generalised to:¹⁴

¹²From Craig (1991).

¹³We only make this assumption for notational simplicity - we could just as easily work with $U(y, B, I; X)$, where I is an indicator for receiving a benefit.

¹⁴Writing the costs as $\sum_{m=1}^M C_m(y; X)I_m$ implies that they are separable: this assumption is not needed.

	Individual: no claim $(U(y + B^*(y; X); X) - C(y; X) \leq U(y; X))$	Individual: claim $(U(y + B^*(y; X); X) - C(y; X) > U(y; X))$
Not entitled ($B^g = B^* = 0$)	No award, not entitled	Never happens
Entitled ($B^g = B^* > 0$)	Would be an award, entitled (genuine non-take-up)	Award, entitled

$$U(y, I_1, I_2, \dots, I_M; X) = \bar{U}(y + \sum_{m=1}^M B_m^*(y; X)I_m) - \sum_{m=1}^M C_m(y; X)I_m$$

where I_m is an indicator variable for taking up benefit m . There are $\frac{M(M+1)}{2}$ possible combinations of benefits, indexed by k . The decision rule is:

$$\text{choose alternative } j \text{ iff } U_j \geq U_k \forall k ,$$

and the observation rule is a generalised form of the table above.

4.2 Generalising the model to allow for expectational and modelling errors

We can expand our discussion to allow for expectation errors on the part of the individual, and errors of the welfare agency (this is based on Duclos, 1995). In particular, we define B^p as the individual's perception of the award were she to apply, ie:

$$B^p = E \{B^g | \Omega^p\}$$

An individual only applies if the perceived award outweighs the costs of applying, or:

$$U(y + B^p(y; X); X) - C(y; X) > U(y; X)$$

(again assuming, for expositional clarity only, separability of $C(y, X)$). Ignoring the analyst's errors, and assuming only 1 benefit, we can now think of an expanded observation rule:

Agency		Individual: no claim ($U(y + B^p(y; X); X) - C(y; X) \leq U(y; X)$)	Individual: claim ($U(y + B^p(y; X); X) - C(y; X) > U(y; X)$)
Not truly entitled ($B^* = 0$)	No award if claim ($B^g = 0$)	Would be no award, not entitled	No award, not entitled
	Award if claim ($B^g > 0$)	Would be an award, not entitled	Award, not entitled
Truly entitled ($B^* > 0$)	No award if claim ($B^g = 0$)	Would be no award, en- titled	No award, entitled
	Award if claim ($B^g > 0$)	Would be an award, en- titled	Award, entitled

Allowing for the analyst's errors as well, we have 16 outcomes:

	Agency	Analyst	Individual: no claim ($U(y + B^p(y; X); X) - C(y; X) \leq U(y; X)$)	Individual: claim ($U(y + B^p(y; X); X) - C(y; X) > U(y; X)$)
$B^* = 0$	$B^g = 0$	$B^a = 0$	Would be no award, not entitled, mod- elled as not entitled	No award, not en- titled, modelled as not entitled
		$B^a > 0$	Would be no award, not entitled, mod- elled as entitled	No award, not en- titled, modelled as entitled
	$B^g > 0$	$B^a = 0$	Would be an award, not entitled, mod- elled as not entitled	Award, not enti- tled, modelled as not entitled
		$B^a > 0$	Would be an award, not entitled, mod- elled as entitled	Award, not enti- tled, modelled as entitled
$B^* > 0$	$B^g = 0$	$B^a = 0$	Would be no award, entitled, modelled as not entitled	No award, entitled, modelled as not en- titled
		$B^a > 0$	Would be no award, entitled, modelled as entitled	No award, entitled, modelled as entitled
	$B^g > 0$	$B^a = 0$	Would be an award, entitled, modelled as not entitled	Award, entitled, modelled as not entitled
		$B^a > 0$	Would be an award, entitled, modelled as entitled	Award, entitled, modelled as entitled

The objects of interest in this model are now $U(y, X)$, $C(y; X)$ (if separable) and the expectations process, $B^p(y; X)$. If we knew B^g, B^a and B^* , we might be able to estimate all of these, because we would be able to place all individuals into 1 of the 16 states above. But there are two problems with this 16-state model (and, to a lesser extent, the 8-state model). First, B^* is never observed. Second, if we observe claims with no award, then B^g is

observable iff someone claims, but if we only observe claims where there is an award, then B^g is not observed unless there has been a successful claim. So, in general, we are not going to be able to distinguish between these 16 states: what appears to be non-take-up due to high stigma costs might look identical - given our data - to non-take-up due to modelling errors or to poor expectations.

4.3 Allowing for endogenous income and take-up

The presentation above has assumed, for convenience, that pre-transfer income is exogenous. This is perhaps reasonable for retired individuals (see Hancock and Barber (2003)). But, if one is concerned with the take-up behaviour of those of working age, then there are two reasons why one should model labour supply behaviour and take-up jointly:

- entitlement to means-tested benefits will generally depend upon labour supply behaviour, and labour supply incentives will be altered by the value of means-tested benefits. This simultaneity implies that, even if preferences for working and claiming benefits are uncorrelated, individuals working and claiming some in-work benefit will have lower propensities to work than those working and not claiming, *ceteris paribus*. Conversely, it means that those observed not working must have relatively high stigma costs, *ceteris paribus*. The group of individuals observed working and claiming an in-work benefit have self-selected themselves into that group, and so inferences based on them may not hold for the population.
- preferences for working and preferences for receiving benefits may be correlated (ie the marginal utilities of leisure, income and stigma may be correlated).

The theoretical specification of this more complex model is almost identical to that above, except the choice set is expanded to include h , hours of work, as well as whether or not to take-up a benefit. For non-random utility models, optimising over this choice set is equivalent to first choosing the optimal labour supply given all of the various combinations of take-up or non-take-up of benefits, and then choosing the combination of benefits that

maximises utility over this restricted set. Observation rules can, in theory, be written down as above, but are more complicated.

4.4 So, what information is needed to model take-up?

Clearly, what information one needs depends on what one wants to know. The list below summarises the models discussed above in increasing order of information demands; all the methods above take the agency's estimate of entitlement to be accurate: none can ever identify agency errors directly because B^* (theoretical entitlement) can never be observed.

1. Simple take-up studies can use data on receipt of a benefit and data on modelled entitlement to model non-take-up conditional on labour supply, assuming no errors by the analyst or the agency, and perfect expectation of benefit entitlement by individuals (data on the amount of recorded receipt is typically not used except as a check that modelled entitlement is not drastically wrong).
2. The development in Duclos (1995) allows the same data to model non-take-up with modelling error (defined as the difference between the agency's and analyst's estimates of entitlement, for whatever reason), but still assuming perfect expectation and exogenous labour supply. This can disentangle apparent non-take-up due to modelling errors from genuine non-take-up, and, if the modelling error process is specified correctly, will eliminate the potential inconsistencies in estimation due to modelling errors.¹⁵
3. Survey data on unsuccessful claims allow the same data to tell us something about expectational error and agency error.
4. Administrative data on incomes and awards from successful claims allows the difference between the agency and the analyst's entitlement estimates to be estimated.

This could be used as a correction factor to some of the estimates above.¹⁶

¹⁵However, Pudney (2003) shows that the likelihood function in Duclos (1995) is not consistent with the presented assumptions.

¹⁶If we believed data from surveys on how much people receive, then we could use that data as a check on the analyst's modelling error.

5. Administrative data from successful and unsuccessful claims, particularly if matched with survey data that contains non-claimants, allows individuals' expectational errors to be modelled, and can potentially disentangle non-take-up due to bad expectations from non-take-up due to stigma.

In most of the economic studies reviewed below, administrative data has not been available, and so researchers have focused on extracting the most information from data on receipt and modelled entitlement from cross-sectional household surveys. As a general rule, panel data always helps the researcher, although the non-linear, discrete choice models do not lend themselves as easily to the conventional fixed effects and random effects models used with linear models (see Honore, 2002).

5 Estimation strategies

This chapter reviews how various studies have sought to estimate models of non-take-up. The general estimation strategy employed by most researchers is to write down an observation rule as we did above, specify a utility function or a function for net utility of receiving a benefit (so some function of $U(y, X)$ and $C(y; X)$) and perhaps the expectations process $B^p(y; X)$, and (usually) use maximum likelihood techniques to estimate the parameters. Below, we give some examples in more detail. Throughout what follows, let I_b be an indicator for receipt of benefit b , which is worth B^* .

5.1 Exogenous labour supply, no modelling or expectational errors

As set out in the previous chapter, our economic model is that families will take up some benefit $B^*(y; X)$ if:

$$U(y + B^*(y; X); X) - C(y; X) > U(y; X)$$

where y is original income, X describes the family characteristics, $C(y, X)$ is the utility cost of claiming and receiving the benefit, which we have assumed is separable, and we assume that families costlessly know their entitlement, and that there are no errors (ie

$B^* = B^g = B^a$) in the terminology used above). If there are some unobservable or random preference or taste effects, then:

$$\Pr(\text{take} - \text{up}) = \Pr [U(y + B^*(y; X); X) - C(y; X) > U(y; X)]$$

Given a functional form for $U(., z)$ and $C(., z)$, and a distribution of unobservable terms across households and assuming $B^* = B^a$, leads to a relatively simple model. For example, Blundell et al (1988) assumes that $U(., z)$ and $C(., z)$ are linear, leading to a “net utility of claiming” function, $V(I, B; X, y)$, defined only on people who are entitled (ie where $B^a > 0$):

$$V(I, B; X, y, B^a > 0) \equiv U(y + B^a; X) - C(y; z) - U(y; X) = \psi(B^a) - (X'\beta - \epsilon) ,$$

where $\psi(B^a)$ is some function of benefit entitlement B^a .¹⁷ Note that this also assumes that unobserved characteristics affect only the cost of claiming benefits: the coefficient on the utility of benefit received is constant across individuals.¹⁸ The observation rule is:

$$\text{observe } I_b = 1 \quad \text{iff } V = \psi(B^a) - X'\beta + \epsilon > 0$$

ie:

$$\text{observe } I_b = 1 \quad \text{iff } \epsilon > -\psi(B^a) + X'\beta$$

Denoting $F()$ as the cumulative distribution function for ϵ :

$$\Pr(I = 1) = 1 - F(-\psi(B^a) + X'\beta)$$

We require that the variables in X - perhaps including family income and composition - are exogenous for a consistent estimator. This is the model used in Blundell et al (1988), Fry and Stark (1993), Riphahn (2001), Hancock and Barber (2003), and Brewer et al (2003a). Non-parametric or semi-parametric techniques could also be appropriate here.¹⁹

¹⁷The costs of claiming the benefit *excluding* the effects of the extra income ($X'\beta - \epsilon$) were not constrained to be positive. This could be seen as a weakness, or it could allow for people receiving the benefit to gain some utility from it that did not vary with entitlement.

¹⁸This is a common assumption, and means that the unobserved characteristics can enter the utility function linearly.

¹⁹These techniques would relax the assumptions about the error terms needed to estimate probit or logit models. See, for example, Pudney (2001).

The net utility function above can be generalised to the multiple benefit case to:²⁰

$$V = \sum_b [I_b \psi_b (B_b) - I_b (X' \beta_b - \epsilon_b)] ,$$

(where b indexes benefits). But estimation of this model by normal ML methods can easily become infeasible because of the need to calculate multiple integrals, and to determine the limits of integration of the necessary integrals, and simulated maximum likelihood techniques would be required.²¹

There are a few cases where estimation is feasible though. In particular, in the unlikely case that there is no financial interaction between the benefits (ie entitlement to one benefit does not depend upon entitlement to another) then the problem reduces to a multinomial probit/logit. Assuming independence of the error terms in the cost functions would also simplify it in this way even if the benefits did interact financially, but it is very likely that there are family-specific unobservable cost-of-claiming factors common to all benefits. Assuming both independence of the error terms in the cost functions and no financial interactions would lead to separate, independent, discrete choice functions.

A two-benefit system, though, does lead to a tractable solution even if the benefits interact, but I know of no examples that use this approach in the literature (although, as described below, Keane and Moffitt (1998), go for the more complicated approach of modelling take-up of three means-tested benefits jointly with labour supply). But this relatively simple model could be used to model the joint take-up of income support and housing benefit, or of WFTC and housing benefit, allowing for that fact that receiving HB reduces the financial incentive to claim WFTC, or (with some modification) of WFTC and the childcare tax credit.

²⁰Again, for simplicity we ignore the restriction that arguably $\epsilon_b \geq -X' \beta_b$.

²¹See Train (2003).

5.2 Exogenous labour supply, modelling errors

5.2.1 Modelling errors and misclassification in discrete choice models

Modelling or misclassification errors in a limited dependent variable model are usually more serious than in a linear regression²². First, probit or logit models are inconsistent when there are misclassification errors, as explored by Pudney (2001). Second, the sample for a model of take-up is defined by the analyst's modelling of entitlement. If entitlement is modelled incorrectly, then this can mean both that some observations are wrongly excluded or included from the sample, and that a variable on the right-hand side is measured with error. For example, Blundell et al (1988) examines whether coefficient estimates are robust to using a semi-parametric estimator which allows for misclassification, and they test for exogeneity in the entitlement variable. But these methods do not account for the fact that the same error in modelling entitlement could affect both the sample used for estimation and a right-hand-side variable. This can only be done by parameterising the modelling error.²³

5.2.2 Parameterising the modelling and misclassification errors

An example of how to model non-take-up with modelling errors with a data-set that does not record claims that lead to no award is given in Duclos (1995, 1997). To simplify his model, it is necessary to assume no uncertainty or expectational errors by individuals, and to condition on observed labour supply behaviour. The way to model modelling errors is to remove the implicit rule that entitlement is non-negative. Most means-tested benefits have a maximum credit and a taper, with negative values implying that the family is not eligible. We reinterpret this approach, calling \overline{B}^* this theoretical benefit entitlement. Clearly, negative benefits are not paid, so $B^* = \max\{0, \overline{B}^*\}$. Costs of take-up are assumed to be $X'\beta + \epsilon$, and the benefit of receiving the money is $\psi(B^*)$, so net benefit (in money terms) to a family deciding whether to claim (and who can anticipate entitlement perfectly)

²²See Hausman (2002) for an introduction.

²³Hancock and Barber (2003) show how measurement error in incomes recorded by a household survey can affect estimates of take-up models.

is:

$$V(B; X) = \begin{cases} -(X'\beta + \epsilon) & \text{if } \overline{B^*} < 0 \\ \psi(B^*) - (X'\beta + \epsilon) & \text{if } \overline{B^*} \geq 0 \end{cases}.$$

(in the notation of above, we have subsumed y into X). A unit will take-up the benefit if $V \geq 0$ (ie only those with positive entitlement and low take-up costs). The log-likelihood is:

$$\log L(\{I\}; X) = \sum_{\{I=1\}} \log P(V > 0) + \sum_{\{I=0\}} \log P(V \leq 0)$$

where I is an indicator for receipt of the benefit. Duclos then assumes that the analyst estimates theoretical entitlement with an additive error: $\overline{B^*} = \overline{B^a} + v$, and that $v \sim N(\mu_v, \sigma_v^2)$ and $\epsilon \sim N(0, \sigma_\epsilon^2)$ and they are independent (the mean analyst error μ_v is allowed to be non-zero, and depend on observed characteristics such as being self-employed). Net utility of claiming a benefit is linear in the amount of benefit received, and non-benefit income is included in the X . Costs of take-up are also constrained to be non-negative ie $X'\beta + \epsilon \geq 0$ (so ϵ is a truncated normal distribution with a different truncation point for each individual). Having done this, it is possible to write down an observational rule, and then derive and maximise the log-likelihood. Identification of the modelling error component comes from the presence of recipients who are modelled as not being entitled, along with the functional form assumption that modelling errors are symmetric and normal.²⁴ The model gives estimates of the determinants of the stigma costs, β , as well as the modelling error and unobserved stigma costs (ie μ_v, σ_v and σ_ϵ).²⁵

It would be possible to pursue this approach with WFTC, assuming labour supply to be exogenous. In the case of WFTC, modelling error would represent the fact that WFTC awards are fixed for 6 months, and this technique might suggest how important this problem was when estimating take-up. It would be appealing to pursue a semi-parametric specification of the modelling errors.

²⁴This “modelling error” represents the difference between the analyst’s and agency’s estimates of entitlement for whatever reason.

²⁵Although, as mentioned above, Pudney (2003) shows that the likelihood function in Duclos (1995) is only correct if there are no errors by the analyst, and that the only source of modelling error is the government agency.

5.3 Exogeneous labour supply, expectational errors

I know of no studies that have estimated expectational errors. As mentioned above, this requires data on successful and unsuccessful claims, ideally from administrative sources and household surveys. There are some possibilities, though

- data that has been used to evaluate the impact of the EMA pilots contains information on unsuccessful claims for EMA (Brewer et al (2001)) .
- data from the new English Longitudinal Survey of Aging will be matched into administrative records, and could tell us about both agency and analyst errors, and expectational errors.

5.4 Endogenous labour supply, no modelling errors

The models in this section have estimated labour supply and take-up of benefits simultaneously. We consider models presented in five studies: Moffitt (1983), Hoynes (1996), Bingley and Walker (1997), Moffit and Keane (1998) and Blundell et al (1999). The principle behind all models remains as above - for the given utility function, specify the observation rules and maximise the likelihood - but the models vary in various technical aspects.

Notationally, we now work with the direct utility function as some function of income, hours of work, benefit entitlement and receipt of means-tested benefit (respectively, y, h, B, I), and (usually) use maximum likelihood techniques to estimate the parameters. Below, we give some examples in more detail. Throughout what follows, let I_b be an indicator for receipt of benefit b , which is worth B^* (we are ignoring modelling errors from now on).

5.4.1 Moffitt (1983)

This study looks at female-headed households and take-up of AFDC. The direct utility function (suppressing characteristics X) is:

$$\log U(y, h, B, I) = -\log(\beta - \delta h) - \frac{\delta(h - \alpha - \delta(y + \gamma B))}{\beta - \delta h} - \phi I .$$

The model allows income from AFDC to be valued differently from other income (through γ , with $\gamma < 1$ implying that benefits were not valued as highly as other income) and for there to be some fixed stigma, ϕ (α, β and δ are other parameters to be estimated). Wages are modelled so that non-workers can be included. Having assumed or averaged away the non-linearities in the tax and welfare system, and assumed additive normal error terms in the (not shown) hours equation and take-up model, the model gives a simple form for h (Tobit) with endogenous take-up. The study finds $\phi > 0$ and $\gamma > 1$, the latter suggesting a mis-specification (such as omitting the value of Food Stamps).

5.4.2 Hoynes (1996)

This study looked at couples with children participating in the AFDC-UP programme. The direct utility function is Stone-Geary, with separable stigma costs:

$$U(y, h_m, h_w, I) = \beta_h \log(\gamma_m - h_m) + \beta_w \log(\gamma_w - h_w) + \beta_C \log(\gamma_C - y) - I\delta$$

where m and w subscript indicates husbands and wives, y is joint income, and c indicates consumption/incomes. Unobserved heterogeneity in distaste for work and claiming AFDC is modelled as a discrete-form approximation, which allows preferences for work and participation to be correlated. This correlation coefficient is estimated to be significantly different from zero, meaning that there two ways that people self-select into AFDC-UP: even if preferences for work and AFDC-UP participation were uncorrelated, those observed on AFDC-UP would be those with below-average preferences for work, but the fact that preferences for work and AFDC-UP participation are correlated means that if AFDC-UP were abolished, the current recipients would still have below-average labour supply.

5.4.3 Bingley and Walker (1997)

This study looked at lone parents in the UK, and the decision to take-up family credit. The utility function as estimated is identified only relative to the utility of non-participation (in other words, they estimate the utility differences between part-time work and non-participation, part-time work, family credit and non-participation, full-time work and non-

participation), which makes direct interpretation of the coefficients difficult. The model allows for involuntary unemployment to prevent the stigma term on family credit from having to explain all observed unemployment/non-participation.

5.4.4 Moffitt and Keane (1998)

This study looked at female-headed households and their decision to claim AFDC, Food Stamps and subsidised housing. The direct utility function is a flexible-form quadratic in its arguments:

$$U(y, h, \{B_b\}, \{I_b\}) = \alpha h + y - \beta_{hh} h^2 - \beta_{yy} y^2 - \sum_{b=1}^3 \Psi_b I_b + \sum_{b=1}^3 \sum_{c>b}^3 \phi_{bc} I_b I_c + \beta_{hy} h y - \sum_{b=1}^3 \delta_b h I_b - \sum_{b=1}^3 \eta_b y I_b .$$

where b indexes the three benefits, and $y = y(\{B_b\}, \{I_b\})$ is income including whatever combination of benefits is claimed. This gives 8 possible combinations of benefit take-up, combined with three choices of hours of work, and they are able to estimate the model by adding an extreme value error term to the direct utility function. Personal characteristics are allowed to affect preferences for work, income and take-up. Identification of the stigma terms arises because some households are not eligible for these benefits.

5.4.5 Blundell et al (2000)

This study looked at all working-age households with children and the decision to claim family credit. The direct utility function was also assumed to be a flexible-form quadratic, as Moffitt and Keane (1998). Stigma was modelled in a less complex way, with the direct utility function being:

$$U(y, h, I) = \alpha_{11} y^2 + \alpha_{22} h^2 + \alpha_{12} h y + \beta_h h + \beta_y y - I \eta$$

where η has some common mean and variance (ie there is no relationship between η and personal characteristics), income includes fixed work costs and childcare costs, and wages are modelled for non-workers. As before, identification of the stigma terms arises because some households are not eligible for these benefits (otherwise it would be indistinguishable

from some general fixed costs of working), and SML is used to estimate the parameters. The choice set is over some set of hours of work, modelled jointly for couples. An almost-identical model is used in Brewer et al (2003).

5.5 Valuing the stigma costs

Any model of non-take-up that directly models the utility function is able to quantify in some way the magnitude of the stigma costs (models where the stigma costs vary with observable characteristics can also value the additional stigma costs arising through changes in observable characteristics).²⁶

For example, if we write a general utility function (not separable in stigma costs) as: $U(y, I; X)$, where y includes B as appropriate and I indicates take-up, then an obvious measure of the stigma costs in utility terms is $C(X) = U(y, I = 0; X) - U(y, I = 1; X)$ at some y .

Then, there are four obvious monetary measures of the stigma cost. Three of these are expressed in terms of income: $y^{CV} - y$, $y - y^{EV}$ and y^* , defined as:

$$\begin{aligned} y^{CV} : U(y^{CV}, I = 1; X) &= U(y, I = 0; X) && \text{at } y \text{ including benefit income} \\ y^{EV} : U(y, I = 1; X) &= U(y^{EV}, I = 0; X) && \text{at } y \text{ excluding benefit income.} \\ y^* &= \frac{C(X)}{\left(\frac{\delta U(y; X)}{\delta y}\right)} && \text{at } y \text{ excluding benefit income.} \end{aligned}$$

As the notation suggests, $y^{CV} - y$ is the compensating variation, $y - y^{EV}$ is the equivalent variation, and y^* is a linear approximation to those two (or the value where stigma costs are marginal).

The fourth measure is in terms of benefit entitlement, and is the level of of entitlement at which families are indifferent between taking up some benefit or not, ie \tilde{B} defined by:

$$\begin{aligned} \tilde{B} : U(y + \tilde{B}, I = 1; X) &= U(y, I = 0; X) && \text{or} \\ \tilde{B} : \Pr(I = 1; y, X) &= \Pr(I = 0; y, X) && \text{at } y \text{ excluding benefit income.}^{27} \end{aligned}$$

²⁶See also Pudney et al (2002).

²⁷ \tilde{B} is identical to $y^{CV} - y$ if benefit income is valued in the same way as other income, but its advantage

The more complex models which directly estimate a utility function (or, in the case of Bingley and Walker (1997), the differences in utility), are able to compute some of the values stated above. But few studies actually report their estimates in great detail, although some information can be inferred from the reported parameter estimates. For example:

- Moffitt (1983) reports the stigma cost of receiving Food Stamps in utility terms, but does not convert into the financial equivalent.
- Moffitt and Keane (1998) directly value in dollars the impact of changes in the explanatory variables on stigma costs (for example, they estimate that a one year increase in age increases the stigma cost (ie y^{CV} above) of Food Stamps by \$1.80 a week), but they do not report the mean stigma cost in money terms of receiving benefits.
- Bingley and Walker (1997) estimate the mean $y^{CV} - y$ for lone mothers for family credit to be £5.91 a week, compared to mean receipt of £25.
- Duclos (1995) reports the implied stigma costs in benefit income terms of receiving Supplementary Benefit for a number of different families in his data-set (ie \tilde{B} above); his model, though, does not estimate a pure utility function, so we cannot express it in terms of non-benefit income. He also finds that unobserved modelling error has a far greater variance than unobserved stigma costs.
- Pudney et al (2002) estimate \tilde{B} for pensioners claiming income support.

6 Conclusions

This paper has summarised the economic literature on modelling take-up, with a focus on income-related or work-related benefits or tax credits. It has given a framework for

is that it is defined by all the take-up models that we have considered above whereas y^{CV} is not. For example, the most straightforward models described above do not estimate a direct utility function, but instead estimate the net utility to claiming some benefit. As the formula suggests, $y^{CV} - y$ requires a direct utility function.

thinking about aggregate measures of take-up rates, and shows how existing studies differ from each other in their data requirements and modelling complexity. It also shows how models of non-take-up, when appropriate data are available, can account for measurement and expectational errors, and how they can value the costs of claiming and receiving some benefits.

References

- [1] Andren, T. (2003), “The Choice of Paid Childcare, Welfare and Labour Supply of Single Mothers”, *Labour Economics*, 10, pp133-147.
- [2] Besley, T. and Coate, S. (1992), “Workfare vs welfare: incentive arguments for work requirements in poverty-alleviation programs”, *American Economic Review* 82, 249-261.
- [3] Bingley, P. and Walker, I. (1997), “The Labour Supply, Unemployment and Participation of Lone Mothers in In-Work Transfer Programmes”, *Economic Journal*, 107, 1375-1390.
- [4] Blundell, R, Duncan, A, McCrae, J and Meghir, C. (1999), “Evaluating In-Work Benefit Reform: the Working Families’ Tax Credit in the UK”, paper presented at the Institute for Poverty Research Conference, North Western University, November 1999.
- [5] Blundell, R, Duncan, A, McCrae, J and Meghir, C. (2000), “The Labour Market Impact of the Working Families’ Tax Credit”, *Fiscal Studies*, 21(1).
- [6] Blundell, R., Fry, V. and Walker, I. (1988), “Modelling the Take-Up of Means-Tested Benefits: The Case of Housing Benefit in the United Kingdom”, *Economic Journal* 98(conference), 58-74.
- [7] Brewer, M., Dearden, L., Emmerson, C., Hayden, C. (2001), “Modelling Take-Up of the Education Maintenance Allowance”, mimeo: IFS.

- [8] Brewer, M., Duncan, A., Shephard, A., and Suárez, M.-J. (2003a), “Did the WFTC work? Analysing the impact of in-work benefits on labour supply and programme participation”, mimeo, IFS, and forthcoming as Inland Revenue Working Paper.
- [9] Brewer, M., Suárez, M.-J., and Walker, I. (2003b), “Modelling Take-Up of Family Credit and the WFTC”, mimeo, IFS, and forthcoming as Inland Revenue Working Paper.
- [10] Craig, P. (1991), “Costs and Benefits: A Review of Research on Take-up of Income-Related Benefits”, *Journal of Social Policy*, Vol. 20, pp 537-566.
- [11] Department of Work and Pensions (2003), ”Income Related Benefits Estimates of Take-Up in 2000/1”, London: Government Statistical Service.
- [12] Dorsett, R. and Heady, C. (1991). “The take-up of means-tested benefits by working families with children”, *Fiscal Studies*, 12(4).
- [13] Duclos, J-Y. (1995), “Modelling the Take-Up of State Support”, *Journal of Public Economics*, 58, 391-415.
- [14] Duclos, J-Y. (1997), “Estimating and Testing a Model of Welfare Participation: the Case of Supplementary Benefits in Britain”, *Economica*, 64, pp81-100.
- [15] Finlayson, L. and Marsh, A. (1998), *Lone Parents on the Margins of Work*, DSS Research Report 80, Leeds: CDS.
- [16] Fry, V. and Stark, G. (1993), *The Take-Up of Means-Tested Benefits, 1984-1990*, London: IFS.
- [17] Hausman, J. (2002), “Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left”, *Journal of Economic Perspectives*.
- [18] Hancock, R. and Barber, G. (2003), “The Quality of Social Security Data in the British Family Resources Survey”, forthcoming, *Journal of the Royal Statistical Society* (Series A: Statistics in Society).

- [19] Hancock, R., Pudney, S., Barker, G., Hernandez, M., and Sutherland, H., (2003), “The Take-Up of Multiple Means-Tested Benefits by British Pensioners”, mimeo.
- [20] Hoynes. H. (1996), “Welfare Transfers in Two-Parent Families: Labour Supply and Welfare Participation Under AFDC-UP”, *Econometrica*, 64(2), 295-332.
- [21] Inland Revenue (2002), *Working Families’ Tax Credit Estimates of Take-up rates in 2000-01*, London: Inland Revenue.
- [22] Keane, M. and Moffitt, R. (1998), “A Structural Model of Multiple Welfare Program Participation and Labour Supply”, *International Economic Review*, 39(3), 553-589.
- [23] McConaghy, M., Hill, C., Kane, C., Lader, D., Costigan, P., and Thornby, M. (2003), *Entitled but not Claiming? Pensioners, the MIG and the Pension Credit*, DWP Research Report 197, Leeds: CDS.
- [24] McKay, S. (2002), *Low/Moderate-income Families in Britain: Work, Working Families’ Tax Credit and childcare in 2000*, DWP Research Report 161, Leeds: CDS.
- [25] McKay, S. (2003), *WFTC in 2001*, DWP Research Report 180, Leeds: CDS.
- [26] Marsh, A., McKay, S., Smith, A., and Stephenson, A. (2001). *Low-Income Families in Britain*, DSS Research Report 138, Leeds: CDS.
- [27] Moffitt, R. (1983), “An Economic Model of Welfare Stigma”, *American Economic Review*, 73(5), 1023-1035.
- [28] Pudney, S. (2001), “The impact of measurement error in probit models of benefit take-up” (mimeo).
- [29] Pudney, S. (2003), “Measurement error in a Structural Model of Welfare Participation: A Comment on a Study By Duclos”, mimeo, University of Leicester.
- [30] Pudney, S., Hernandez, M. and Hancock, R. (2002), “The Welfare Cost of Means-Testing: Pensioner Participation in Income Support”, paper presented at the Royal Economic Society Conference, Warwick 2003.

- [31] Remler, D.H., Rachlin, J.E. and Glied, S.A. (2001), "What Can The Take-Up of Other Programs Teach Us About How to Improve Take-Up of Health Insurance Programs?", NBER Working Paper 8185.
- [32] Riphahn, R.T. (2001), "Rational Poverty or Poor Rationality? The Take-Up of Social Assistance Benefits", *The Review of Income and Wealth*, Volume 47(3), pp379.
- [33] Train, K. E. (2003), *Discrete Choice Methods With Simulation*, Cambridge: CUP.
- [34] Wicks, R. (2003), *Challenging take-up: means-testing and tax credits*, London: SMF.
- [35] Yaniv, G. (1997), "Welfare Fraud and Welfare Stigma", *Journal of Economic Psychology*, 18, 435-451.