

**How Important is Income in
Determining Children's
Outcomes? – A Methodology
Review of Econometric
Approaches**

**Laura Blow,
Alissa Goodman,
Greg Kaplan,
Ian Walker,
Frank Windmeijer**

1. Introduction

“Children who grow up in poverty experience disadvantage that affects not only their own childhood, but also their experience as adults and the life chances of their own children. Support for today's disadvantaged children will therefore help to ensure a more flexible economy tomorrow.”

- HM-Treasury Budget Report 2003

Policymakers have used the wide body of research evidence from around the world, which purports to show that children who grow up in poor families experience a wide range of negative outcomes, to justify large increases in benefits to families with children.

Since 1997, for example, the British government has increased the level of both universal and means-tested benefits, and tax credits payable to families with children by a very considerable amount. For example, Adam and Brewer (2004) show that government spending on child-related benefits for families with children has increased by 52% in real terms between 1999 and 2003, and child-contingent support payments now account for 2 per cent of GDP.

The most commonly found justification for this approach has been that it serves to improve children's outcomes, both in childhood and later in life (see for example HM Treasury, 2001).

But we have remarkably little evidence as to how much children's outcomes are likely to improve as a result of giving additional money to their families, and even less

on how effective this is likely to be in comparison to spending the same public money in other ways, for example on improving school quality, or neighbourhood services.

The evidence that we do have - mostly from the US, but some also from the UK and elsewhere – shows that income supplementation may not be a particularly effective means of improving children’s outcomes¹. How its effectiveness compares to other policy tools available is much more of an open question.

Our lack of evidence on these important issues stems from the fact that researchers have struggled to find effective methodologies to identify the separate impact of low income in itself on children, from the other factors – both genetic and environmental - impacting on children from low income households.

This document reviews a number of empirical methodologies available to researchers for answering these important policy questions. For each methodology we set out the intuition behind the method, and an example of where it has been used effectively to answer a relevant policy question. We restrict ourselves to methodologies found within the econometric literature.

This work forms the first stage of a larger research programme in which we attempt to apply some of the methods described here to give robust answers to policymakers as to the likely effectiveness of income supplementation policies.

¹ This does not mean that such income supplementation is not justified; simply that justification for it cannot rest solely on the grounds of its impact on children’s outcomes.

We start by setting out the circumstances in which income could matter for determining children's outcomes, before addressing the issue of how such an effect could be measured.

2. Specifying the model

The workhorse of econometric analysis of survey data is the linear regression model. In order to fix ideas, consider the following simple linear regression model of the determination of an outcome Y_i for a child who grew up in family i with income M_i

$$Y_i = X_i'\beta + \lambda M_i + u_i \quad (2.1)$$

Here X represents other control variables that affect the outcome Y and u is a residual or error term that includes omitted determinants of Y including measurement error. The challenge is to specify and estimate this model to ensure that the causal effect of income on the outcome, λ , is correctly recovered. Even this simple model can be used to illustrate the key issues for the researcher to consider in meeting this challenge. These issues include:

i) Deciding how income, M , enters the model and what type of income matters. In our simple illustration income enters linearly, but it could be quadratic or logarithmic or enter in some other way. In addition our model assumes the income effect is the same for all households (a homogenous effect) whereas it could differ across households (a heterogeneous effect) which we would indicate by letting λ have an i subscript.

ii) *What else matters?* Not only do we need to specify correctly how income should enter the model, but we also need to decide what other characteristics, X , should also enter the model. We want to include covariates X that simultaneously affect income, M , and the outcome, Y . For example, suppose parental education has a positive effect both on parental incomes and on child outcomes. Then, if we omit parental education from our empirical specification, we will overstate the causal effect of income. This effect is often called omitted variable bias. In the extreme, we could wrongly ascribe a causal effect to income where there is none, because income happens to be strongly related to an omitted variable that has the true causal effect.

iii) *What methods of estimation are appropriate?* One of the key assumptions in equation (2.1) is that the conditional expectation of the error term is zero, $E[u_i | X_i, M_i] = 0$. If this is not the case then estimation of the model by standard simple technique of ordinary least squares (OLS) will result in biased coefficients. There are many reasons why this assumption might fail, and the general problem is usually referred to as endogeneity. Indeed point ii) above can be thought of as an illustration of endogeneity since omitted variable bias results from M being correlated with a variable that has been relegated to the error term. However this can be overcome by correctly including all the relevant observable covariates. A bigger challenge in empirical work is dealing with the possibility that there are unobservable (to the researcher) parental characteristics that affect both income and parental abilities. For example, parents have genetic endowments such as natural ability or good health that may help them to earn

a good income *and* to be a successful parent. Moreover, these endowments may well be inherited by their children, and have a positive effect on their life outcomes. In this example unobservable factors that contribute to better child outcomes are also *positively* correlated with parental incomes and so OLS estimates of the impact of income would be biased *upwards*. Thus least squares is still useful because it at least gives an upper bound to the effect of M . Unfortunately, this is not a generalisable result, since we cannot assume that M and u are always positively correlated. We should not only be concerned by the possible endogeneity of income, but also of the covariates, X . Even if we do not care about estimating unbiased coefficients, β , for the effects of these variables, including them in the model without taking account of possible endogeneity is very likely to bias our estimate of the causal effect of income, λ . The general point is that in the presence of endogeneity OLS is not the correct estimation technique and the researcher needs to decide how the model can best be estimated, dealing with the confounding effects of any unobservable factors, u , that affect both parental incomes (or covariates) and child outcomes.

We briefly consider these issues further here, before going on to discuss possible estimation techniques in Chapter 3.

2.1 Why and how should income matter?

Why should income matter?

We have said that choosing the correct functional form for income in equation (2.1) is paramount, but we have not discussed how this should be decided. Whilst it is possible for the researcher to try and “let the data do the talking”, it is preferable to have a theory in mind, since this will provide guidance on how the causal effect of income occurs (and, for that matter, on what other variables should enter the model and which variables might be endogenous). It is theory that provides us with the framework within which we interpret empirical findings.

An economic theory would be based on some kind of optimising behaviour on the part of individuals or households. Economists take it for granted that by giving additional income to an individual, their welfare will be improved. But understanding how important giving additional income to parents is likely to be for the well being of children is more complex than this. This is because children depend on the behaviours of, and decisions made by, their parents to determine how much, and in what way, they will benefit from additional income into the household.

Theories of parental investment in their children have been the focus of many economists’ thinking about the role of parental income in determining children’s outcomes (see Becker and Tomes, 1986). In the Becker-Tomes model parents care about their child’s future financial wellbeing, which they can affect by investing in the child’s human capital (education and so on). Optimising behaviour implies that they will invest optimally, i.e. to the point where the marginal benefit equals the marginal return. This implies that, if parents are not credit constrained, parental income itself should not matter for investments in children. However, if

parents are not allowed to borrow against their children's future earnings (and it seems unlikely that they would be able to do this, particularly when the child is young) then poorer parents will not be optimally invest in their children and parental income and child earnings become causally linked.

Whether credit constraints are important for families with older children, for example in preventing desirable investments in post compulsory schooling, is more open to debate (see for example, Cameron and Heckman, 2001, and Carneiro and Heckman, 2002 who argue against the presence of widespread liquidity constraints in post-secondary schooling).

Alternatively, if lack of income places a strain on parents and thereby reduces the quality of their parenting, then additional income could make for better children's outcomes, regardless of what the extra money is used to buy. Theories such as the so-called "parental stress" and "role model" theories have tended to emphasise these alternative pathways between changes in household income and children's outcomes (see Mayer, 1997 chapter 3 for a good discussion of this).

Current income, permanent income, or something else?

Though it is current parental income, measured at a snapshot in time, that is most readily available to researchers for assessing the effect of parental resources on child outcomes, this is unlikely to be the most appropriate measure for these purposes. If parents can save, borrow, and run down assets for the benefit of their children, then unless parents are completely

myopic, past values of income (and expected future values) are likely to be important as well as current income. Similarly if there are time lags between parental investments and their impact on child outcomes, then it will be important to include past values of parental income in the model for this reason too.

Many researchers, if they want to avoid simply adopting a measure of current income, use an estimate of "permanent" income – which is usually modelled as income averaged over all periods for which data exist - as a measure of longer run resources available to a family. However, this specification of income could also be inappropriate, since there is no presumption that income from different periods would have the same effect on current outcomes, as would be implied by this commonly used specification.

In practice, this suggests that researchers need to test various specifications of how present and past (if available) values of income enter into their model in order to discriminate between alternative theoretical ideas.

Is the income effect linear?

Many researchers estimate the impact of income on child outcomes, assuming that this effect is a linear one – i.e. that each additional £1 of parent's income results in an equal addition to the child's outcome, regardless of the level of income. However, there is no reason why this should be the case. For example, there may well be decreasing marginal returns with respect to income in

the production of child outcomes. Depending on parental preferences, this could result in a non-linear (concave) income effect. Some researchers use other specifications of the income effect, for example one based on an "income-to-needs" ratio or some other equivalisation procedure. However, there is no clear reason to assume a priori that these specifications will be correct.

This suggests that it is important to test various specifications of the income effect for their validity. More generally, our uncertainty about what the appropriate specification might be suggests that non-parametric methods might usefully be used to model the relationship between outcomes and incomes. However, such methods require considerable amounts of data to be used successfully and also require that income be a continuous variable – something that is not always true in the data available.

Income from different sources

Parents derive income from a number of different sources, and these may have different impacts on children's outcomes. For example, parents may not pool their incomes and might have different views about how their own resources should be allocated that, at the margin, implies a differential impact of mother's income relative to father's income. Moreover, it may be appropriate to treat labour and non-labour income differently when estimating the impact of parental income on children's outcomes. This may be because additional income from work - if obtained through working longer hours - could reduce the amount of time

a parent can spend with their child. Unless this effect on time allocation is controlled for, additional income from work might have a different effect on children's outcomes than additional income from other sources.

Is relative income important?

Relative income levels, rather than absolute incomes, are likely to be important for outcomes that are driven by a rank. For example, suppose that absolute income is very important for determining exam performance that is used as a criterion for university entrance. If university entry is rationed and allocated to the top x% of a cohort then the probability that an individual enters university will depend on own performance *relative* to that of others in the same cohort. Hence *relative* parental resources might matter for University entrance even if it does not for the examination performance that drives it. Moreover if the weight that parents attach to the outcomes for their child is formed with reference to their peers, neighbours or family then relative resources will matter for the outcomes.

Indeed, child development is influenced by many factors and while the most familiar are family factors, such as family composition, and socio-economic status, "neighbourhood effects" are also likely to be important. The idea that the composition and quality of a neighbourhood can affect children independently of family effects might also be captured by income relative to neighbours.

Heterogeneous effects

The effect of income on child outcomes may differ across households for a variety of reasons according to the household's observable characteristics. For example, the effect may be different for highly educated parents than for less educated parents. When causal effects differ according to observable exogenous characteristics then one can simply estimate separate statistical models for each subgroup which allows the income effect and the effect of all other variables included to differ across subgroups – or some more restrictive specification, say one that just allows the effect of the income variable to have different effects across subgroups through interacting income with those characteristics.

When the effect of income differs across the population by unobservable characteristics we cannot split the data into the relevant subgroups. If income is exogenous then simple regression methods estimate the *average effect of income*. However if income is endogenous, in the sense that unobservable factors that affect the outcome also affect income, the problem is more complex. We discuss this further in section 3.3.

2.2 What else should enter the model?

As well as making sure that income enters the model correctly, it is important to include other observable variables (denoted by X above) that determine both child outcomes and parental income. If any factors are excluded that determine child outcomes, and are also correlated to parental incomes, the estimated income effects will be biased.

However it can be difficult to decide exactly which observable variables should be included in X . We *do* want to include variables that affect both parental incomes and child outcomes (we used the example of parental education when illustrating point ii) above), but we *do not* want to include variables that are solely pathways through which parental income exerts its causal influence on child outcomes. As an illustration of the latter, suppose one way in which income affects child outcomes (for example educational achievement) is through the purchase of books for the child. An increase in income will increase book purchases, which in turn affects educational achievement. If we include book purchases in our regression equation then we will get estimates of income effects on educational achievement that are inappropriate for policy analysis, because they hold constant a variable (i.e. book purchases) that would change in response to a policy induced change in income.

A further complication arises since some of the covariates that, in principle, we do want to include may be endogenous. Even if we do not care about the unbiasedness of the coefficients we estimate on endogenous covariates, including them may well bias our estimate of the causal effect of income. For example, we know that part of the variation in income across households in survey data arises from differences in hours of paid market work supplied by the parents. Now suppose that the time that parents spend with their children is important for child outcomes and that harder working parents do not have as much time to spend with their children. Then if we do not include hours worked or time spent with children we will underestimate the

causal effect of income (the omitted variable bias problem). But, it could also be that working hard is partly caused by an unobserved characteristic (call it determinedness) that also makes the parent more effective during whatever time he/she does manage to spend with the child so that hours of work is an endogenous variable. There is no simple solution to this dilemma. In effect we have two endogenous variables, income and hours of work.²

In view of this it is probably important to test the sensitivity of income effects to the inclusion of variables such as education, labour supplies, marital status, fertility, and household composition. For example, if when we introduce hours of market work into the model we find that the positive effect of income on the child outcome (we presume that this is a “good” outcome) becomes substantially larger then we can presume that the time required to earn that income is also indirectly important for the outcome.

² Note that here we are discussing survey, or non-experimental, data. Our reasoning might change in the context of an experiment that randomly assigned extra income to some families (the treatment group) and not to others (the control group). In this case the response to the extra income might be to reduce hours worked, which would also decrease earned income, both of which might have an effect on the child outcome. Here, these responses are directly caused by the income supplement and so are part of the causal effect and we would not want to try and control for them. Indeed, if the experiment was properly constructed we could calculate the mean effect of the income supplement by simple taking the mean difference in outcomes across the treatment and control groups. We discuss experimental data further in section 3.2.

3. Estimation techniques

As we discussed above, the main problem researchers must contend with in finding suitable estimation techniques is that parents' income may be correlated with the disturbance term, u in the child-outcome equation. This term will contain the effect of all unobserved factors – at the level of the child, the family, the local area, the point in time, etc. – that are correlated with child outcomes. If any of these are also correlated with parental income then the estimated effect of parental income on child outcomes in this model will be biased.

The basic ideas that have been employed to overcome this problem is either to:

- Search for *exogenous* variations in income or, at least, exogenous variation in some components of income – that is, variation in income that is not correlated with unobserved parental characteristics;³

or to

- *Difference* out the effect of unobserved characteristics (generally on the assumption that they are fixed over time). There are two possible ways of doing this. One could compare the effect of variations in income over time with the

³ As will be discussed in later sections, exogenous variation that will often be utilised to obtain causal effects is due to policy changes. The estimation methods discussed here are therefore very similar to those in the literature on programme evaluation. It is not our aim, however, to summarise the programme evaluation literature concisely. This is done in for example Blundell and Costa Dias (2000).

variation in outcomes across siblings within the same households. Or one could use similar groups of households, over time, and see how the average outcomes across the children between the groups varies with the average income of the groups.

The extent to which these techniques can be used is, of course, dictated by the data that are available. Generally speaking, the richer the data we have available, the less restrictive are the assumptions we have to make to identify a causal effect. In many cases the data will allow us to employ more than one technique and the resulting estimates will be closely related to each other.

One important issue that we will explain later is that the parameters identified in the different methods have different interpretations. Some are more useful than others for policy purposes.

We start, however, with a discussion of the standard techniques for estimation of the parameters in a linear model using non-experimental survey data, Ordinary Least Squares and Instrumental Variables.

3.1 Ordinary Least Squares and Instrumental Variables

The main expository model outlined earlier considers a simple linear model for the relationship between a child's outcome Y and parental income M :

$$Y_i = X_i'\beta + \lambda M_i + u_i \quad (3.1),$$

where the vector X contains other observed characteristics of the child and its parents. Unobserved characteristics of both the child and its environment are contained in the error term u . We assume here that there is a constant effect of income on child outcomes across the households.

The standard procedure for estimating the parameters in this simple linear regression model is that of Ordinary Least Squares (OLS). This procedure obtains unbiased estimates of β and λ , which measures the effect of parental income M on child outcome Y , only if the unobserved characteristics contained in the error term u are not correlated with income M and the other regressors X .

As we discussed in Section 2, it is very likely that income M is correlated with the unobservables or omitted variables in u . One common approach to this *endogeneity* problem is to find so-called instruments for parental income. These instruments, denoted Z , are observable variables that contain information that is correlated with parental income M (strictly speaking once the effects of the other included exogenous variables, X , have been netted out), but not with the unobservables contained in u . This means that the instruments do not contribute to the determination of the outcome Y , given that the model already includes X and M . If such instruments exist, consistent⁴ parameter estimates can be obtained by means of the Instrumental Variables (IV) estimation technique.

⁴ Parameter estimates are said to be consistent if any bias that they exhibit tends towards zero as the samples becomes large.

Roughly speaking, a first stage regression of M on the instruments, Z , and X produces a prediction for M which no longer involves the part that is correlated with u , and which can then be used in the main regression. Intuitively, Z allows us to isolate exogenous variation in M , i.e. variation that is independent of u .

We can get consistent estimates of the M equation by OLS. So we can compute the predictions of M , \hat{M} . M and \hat{M} are correlated with each other – indeed they are the same apart from random unobservable differences (including the u 's).

The Instrumental Variables estimator is then the OLS estimator of model (3.1) where M is replaced by \hat{M} . Note that Z will have to include instruments that are different from X – if they were the same then \hat{M} would simply be a linear combination of the *same* variables as are already in the Y equation – i.e. \hat{M} would be perfectly collinear with the X 's.

Thus the main problem of implementing IV is the need to find some variable that belongs in Z that does not also belong in X - an “exclusion restriction” is required.

Ideally one would like to be able to test the assumption that the instruments are not correlated with the error term, but this is not possible when there is one instrument for the potentially endogenous variable (this case is called “just identified”), since the estimation technique gives residuals that are, by construction, uncorrelated with the instrument. When the equation is over-identified (more instruments than endogenous variables) a test of the “over-identifying restrictions”

can be constructed which gives an indication whether the instruments are valid in the sense that they are not correlated with the unobservables in the equation of interest.

Bias in IV

The consistency of IV places strict requirements on the instruments – they need to be correlated with M and not with u . “Weak” instruments (Z not well correlated with M) and instrument invalidity (Z correlated with u) can lead IV to be even more biased than OLS.

If we have more than one exclusion restriction (ie more than one IV) then we could test the validity of excluding each. But in “just identified” models (i.e. with only one IV) we cannot easily test the strength or validity of the IV. Bound et al. (1995) demonstrate the extent of bias in IV and suggest rules of thumb for instrument validity and strength in the just identified case.

The existence of valid instrument(s) is even necessary to test whether there is an endogeneity problem – an appropriate procedure is to test for endogeneity by including the estimated residual from the instrumenting equation into the child outcome equation – if it is statistically significant then there is evidence of endogeneity.

Heterogeneous effects in IV

Conventional applications of the IV method assume that the effect of the variable of interest (parental income in our case) has the same effect on the outcome across

households (or, at least, for all individuals with the same value of other X 's in the model).

When the effects of income differ according to a variable like education that is itself correlated with the outcome, then it becomes difficult to unravel the direct and indirect effects – how much of the better outcomes arise because parents are richer and how much because richer parents are better educated. Since parental education and parental incomes are likely to be correlated we require at least two instrumental variables, one for each endogenous variable.

Another difficulty arises when the effect of income differs across the population for reasons that are unobservable. This problem is discussed further in section 3.3.

3.2 Randomised Experiments

The type of information required to exploit exogenous variation in parental income in any given observational data source can be very difficult to obtain. A way to circumvent this problem is to create such exogenous variation in parental income between children, by means of an experiment that increases parental income or earnings for some children, but not for others.

Although randomized controlled experiments in economics are rare they are nonetheless interesting. Firstly, experiments are commonly thought of as providing the “gold standard” of evaluation so we need to understand how such experiments work. Secondly, partly because of their high reputation, when

experiments are actually conducted their results can be very influential, so it is important to understand their limitations and the factors that can undermine their validity. Thirdly, events sometimes occur “naturally” to produce what seems like randomisation; that is the allocation of M to individuals turns out to be random. For example, suppose that a policy is enacted in one area but not its neighbouring area. If the residents of one area are, on average, identical to those of the neighbouring area then the individuals subject to the policy change can be viewed as a treatment group while the others can be thought of as the control group – because, on average, the two groups are identical. Thus the policy produces a “quasi-experiment” or “natural experiment,” and many of the issues relevant to actual experiments also apply to natural experiments.

In the context of discovering the effect of parental incomes on child outcomes an ideal experiment might be designed in such a way that families who are eligible for some income subsidy are randomly allocated into two groups, one of which actually gets the subsidy (the treatment group) and one of which does not (the control group). For example this might be done by a lottery. As the subsidy is assigned in a completely random manner for this group of families, the distribution of observed and unobserved characteristics of the two groups will be identical. In this case, the income variation induced between the treatment and control groups through the income subsidy will also be completely random with respect to parental and child characteristics, and therefore will not be correlated to any unobserved factors, which could also determine child outcomes, Y .

This makes the effect of the income subsidy on child outcomes very simple to estimate – since the average difference in children’s outcomes between the treatment and control groups represents the average impact of the income subsidy for those that were eligible for it. This difference is written as:

$$\bar{Y}^T - \bar{Y}^C$$

where \bar{Y}^T is the average outcome for the treated, and \bar{Y}^C is the average outcome for the controls. The effect of family income on the children’s outcome measure can be obtained using the simple estimator

$$\hat{\lambda} = \frac{\bar{Y}^T - \bar{Y}^C}{\bar{M}^T - \bar{M}^C}$$

where \bar{M}^T is the average income for the treated, and \bar{M}^C is the average income for the controls and the nature of the lump-sum subsidy implies that their difference is the same for everyone.

If the treatment effect was thought to differ across individuals according to their values of X , for example by race, then one might estimate separate equations for each ethnic group and then compute the treatment effect from the difference in predicted values for each group. In any event, the random assignment to the income subsidy creates exogenous variation in income and this estimator can therefore also be seen as an instrumental variable estimator in the model

$$Y_i = \alpha_0 + \lambda M_i + u_i$$

using D_i as an instrument for M_i , where as $D_i = 1$ if a family received the income subsidy, and $D_i = 0$ otherwise.

In effect an experiment that randomly assigns income subsidies across individuals provides the data to estimate the effect of such a policy from a simple bivariate regression. In the case where the variation in M is bimodal – so that the experiment randomly assigns a specific additional income to the treatment group then λ would be the estimate of *that* change in income – extrapolating from this would then depend on some assumed functional form of the relationship between Y and M . Strictly speaking an experiment tells us *only* about the effects in the experiment – and tells us *nothing* about any other change unless further functional form assumptions are met. If the variation in M is discrete but multivalued so that the size of the treatment differed across the treatment sample then one could adopt some functional form, say linearity. Studies that use multiple treatment groups, where the treatment varies across groups, are known as “differential impact” designs⁵.

It will be instructive to introduce here some terminology that is used commonly in the programme evaluation literature. As we mentioned previously, the estimation

⁵ An example is SSP in Canada where one group were given a financial incentives package while another was given this plus job search advice. The Educational Maintenance Allowances evaluation in the UK had multiple pilot groups, some with different levels of the financial treatment, which could be compared with the matched areas and with each other. The benefit of having such multiple treatments in the single evaluation study is that it allows one to consider each relative to the control group and relative to each other.

methods described here draw heavily upon those that have been developed in the programme evaluation literature.

The average income effects of interest due to an (natural) experiment can be classified as Average Treatment Effects (ATE), the Average Treatment on the Treated (ATT) effects, the Average Treatment on the Non-Treated (ATNT) effects and the Local Average Treatment Effects (LATE), the latter described in detail section 3.3.

ATE is the average child outcome effect in the population due to an exogenous income change. This average effect is the effect for a household randomly drawn from the population and given an exogenous income change.

ATT is the average child outcome effect for households that did get an exogenous income change, whereas ATNT is the average child outcome effect of an income change for those household that did not receive it.

The identification and thus interpretation of a treatment effect is related to some basic assumptions made in the model and what estimation strategy is being employed, both of course dependent on the information and data available to evaluate a treatment effect. For example, if the treatment effect is assumed to be constant, then ATE, ATT and ATNT are all the same. If the income affects child outcomes heterogeneously, ATE, ATT and ATNT are in general not the same, and it is important to understand what is being identified under what assumptions and estimation techniques.

For example in the income subsidy experiment, if there is randomisation into treatment and control groups amongst those that were eligible and did apply for the subsidy, the effect that will be estimated by $\hat{\lambda}$ is the ATT if the effects are heterogeneous in the population.

Potential problems with randomised experiments

Randomised experiments are often taken to be a gold standard for establishing the effects of policy on individuals. But it is important to bear in mind the possible limitations of randomised controlled experiments. It is very important, for example, to ensure that randomisation is implemented correctly, so that the treated and untreated populations have the same distribution of observable and unobservable characteristics.

Beyond this, it is also important that subsequent to randomisation, all people included in the treatment group do indeed receive the treatment (in our case an income subsidy) and do not fail to take it up. Researchers can also encounter problems if there is significant or unbalanced attrition from surveys designed to measure the impact of the treatment. Additionally, as in medical experiments, the mere fact that the subjects are in an experiment can change their behaviour, and this will also tend to bias any evaluation results. All such issues relate to the so-called “internal validity” of any evaluation results. These, and other practical issues associated with social experiments are considered further in Stafford, Greenberg and Davis, 2002.

Problems related to the “external validity” of evaluation results from randomised experiments can also commonly arise. These compromise the ability to generalize the results of an experiment to other populations and settings. Some common threats to external validity arise when the experimental sample is not representative of the population of interest, or when the treatment or policy being studied is not representative of the treatment that would be implemented more broadly. For example, a tightly controlled and carefully monitored experiment could be quite different from the programme actually implemented, if it is not subject to the same quality control as the experimental version. Another difference between an experimental programme and an actual programme could be in its duration: the experimental program only lasts for the length of the experiment, while the actual program under consideration might be available for longer periods of time. A temporary experiment may give rise to quite different incentives than a permanent policy.

An issue related to scale and duration concerns what are called “general equilibrium” effects. Turning a small, temporary experimental programme into a widespread, permanent program might change the economic environment sufficiently that the results from the experiment cannot be generalized. For example, a widespread educational reform, such as school vouchers or reducing class sizes, could increase the demand for teachers and change the type of person who is attracted to teaching, so the eventual net effect of the widespread reform would reflect these induced changes in school personnel. An internally valid small experiment might correctly measure a causal effect, holding constant the

market or policy environment, but general equilibrium effects mean that these other factors are not, in fact, held constant when the program is implemented more widely.

Another potential threat to external validity arises because participation in an actual (non-experimental) programme is usually voluntary. Thus, an experimental study that measures the effect of the program on randomly selected members of the population will not, in general, provide an unbiased estimator of the program effect when the recipients of the actual implemented program are permitted to decide whether or not to participate. A job training programme might be quite effective for the few who choose to take it, yet be relatively ineffective for a randomly selected member of the population. One way to address this issue is to design the experiment so that it mimics as closely as possible the real-world program that would be implemented.

Thus, in general, there are many reasons why real experiments, however well designed, might not provide the information required.

3.3 Natural Experiments

Even where it is not possible to set up an explicitly randomised experiment in order to create exogenous income variation between families, researchers may be able to exploit events that effectively mimic an experiment of the kind discussed in the section above. This approach also aims to exploit exogenous income

variation between families. The same criticisms apply to this type of experiment as to real experiments.

In the context of this review, a useful natural experiment would be an event that affects the parental income of one group of children differently from its effect on another group, in a way unrelated to the unobserved characteristics of the children or parents that determine child outcomes. This could be a policy change - for example a change in child benefit or other state benefit that affects some groups of children differently from others.

Just as a randomly assigned income supplement can be thought of as providing an instrumental variable for M to provide an estimate of the causal impact of M on Y , so too can an income supplement that was naturally assigned providing the resulting assignment was uncorrelated with Y .

The idea is then to identify a causal impact of the income change by comparing children's outcomes of the group that was not affected by the policy change to those of the group that was affected by the policy change. There are various ways of evaluating the policy's impact, which depend, in part, on what type of data is available, and again on whether one thinks the income effects are homogeneous or heterogeneous. We mentioned in section 3.1 that when income effects are heterogeneous, then applying IV methods to overcome endogeneity becomes quite complex. Since IV methods are frequently used in the context of natural policy experiments, we now discuss this in some detail.

Instrumental Variables in the Context of Natural Experiments: Local Average Treatment Effects (LATE)

An extreme example helps to illustrate the issues surrounding the use of IV-type estimators when “treatment” effects are heterogeneous. Suppose that the natural experiment arises because of an increase in the statutory school leaving age. The instrument Z_i records whether an individual faced a minimum school leaving age of 15 or 16 and M_i is parental income. Then, if raising the minimum school leaving age *only* affects the education and incomes of those who wanted to leave at 15 (see Harmon et al. (2003) for evidence of this), an IV estimate of the effects of parental income on the child’s outcome will be a consistent estimate of the effect for those parents that wished to leave school at 15 and for *no-one* else.

Thus under heterogeneity of income effects IV produces consistent estimates – but it is an estimate of the effect on a subgroup of the population that is generally hard to determine. Imbens and Angrist (1994) term this estimator the *Local Average Treatment Effect (LATE)*. In the extreme example above we have some grounds for believing that the reform that forms the basis of the instrument only affected the bottom of the education distribution but in most cases it will be difficult to be specific about the weights that determine the composition of the relevant *local* group. Thus, while IV may provide a consistent estimate it may not be clear who it is an estimate for. If we wanted to know the effect of income for the whole population, perhaps because a prospective policy would increase income across all parents, then having an estimate that is relevant to, say just the poorest parents, may not be very

helpful. But if policy is targeted on the poorest parents then our estimate would be helpful.

3.4 Matching methods

The impact of policy changes is also frequently evaluated using matching methods.

The purpose of experiments is to facilitate the construction of the missing counterfactual by having a control group to tell us what would have happened in the absence of the experiment, or, in the case of a natural experiment, in the absence of the reform.

However when we are considering natural experiments rather than those that are randomised by design, policy reforms often give rise to treatment and control groups that may be quite different from each other in terms of their observable characteristics. This then begs the question of how to control for these observable differences.

If only cross-sectional information is available, a simple estimator of the effect of the policy change on child outcomes is the mean difference between the group affected by the reform and the group unaffected. However, as the two groups may in this case be very different in observed characteristics, such a simple comparison is likely to confound the effects of the reform with the effects of the differences between the groups. In order to compare like with like, a *matching estimator* can be used to compare the outcomes across groups of individuals that are very similar.

Thus, one might group the data into cells – say by race, gender and age group – and take the difference between the means of the treated and untreated observations within each cell. Of course, if there are a large number of characteristics that differ across the two groups then there will be many cells and the number of observations in each cell may become quite small so the precision of our estimate would be low. Moreover, the treated and untreated observations may differ in income or years of education or some other variable that is continuous or has many cells. That treated and untreated observations might be very different in their observable characteristics is known as the *common support* problem and it implies that effective matching methods may become difficult to implement in practice.

Notice that matching here is done using observable characteristics only. Implicit in this that the allocation of the treatment is independent of variables that are unobservable. Thus, it is sometimes argued that matching methods require rich datasets that contain many observables. The difficulty in exploiting the richness of extensive datasets is that so many variables are available for categorising and grouping the data that the data gets grouped into a large number of cells each containing few observations. Indeed, it could be that some cells might contain a very unbalanced distribution of the treated vs the untreated – the common support problem mentioned above. Not surprisingly if we exploit the richness of extensive data to enable us to compare like with like we may end up having few comparable observations and so correspondingly imprecise estimates.

It has been suggested that we can overcome this problem using *propensity score matching*. In this method we model the receipt of the treatment (or, the availability of the treatment if we were considering some policy where individuals have to elect to receive) as a function of observable variables and predict the probability of treatment for both the treated and untreated groups. Propensity score matching then compares the outcomes across these two groups between individuals that have a very similar probability of receiving the treatment. In order to do this, the researcher will first estimate a probit or logit model for this probability:

$$p(D_i = 1|x_i)$$

estimated on the sample of treated and controls. This is then used to obtain the estimated probabilities \hat{p}_i . For every \hat{p}_i^T , i.e. the estimated probability for a treated person, the researcher then finds the closest \hat{p}_j^C of the controls (either by choosing the nearest neighbour, or some combination of controls that matches each treated individual). The propensity score matching estimator for the treatment effect is then given by

$$\frac{1}{N_T} \sum_{i=1}^{N_T} (Y_i^T - Y_j^C)$$

Note that no functional form specification for the relation between the outcome variable and any other characteristics has been specified when deriving the policy impact. This makes the specification completely flexible.

Matching here is being done, not on X but on $p(D_i = 1|x_i)$. Nevertheless, for matching to be successful, there are two important requirements. First, there must still be *common support* - i.e. enough individuals with similar probabilities of being treated. Second, it is also extremely important that individuals in treatment and control groups only differ with respect to characteristics observable to the researcher.

If the distribution of unobserved characteristics (e.g. parental ability) differs then estimates of the effect of the policy will be biased. However whether or not this strong requirement is met can – by definition – never be tested. For this reason it is very important that the natural experiment chosen divides individuals into treatments and controls in a way that means that the groups are unlikely to differ by their unobserved characteristics.

3.5 Control function approach

If the treatment and control groups are believed to differ, after controlling for observable characteristics, in unobservable characteristics that determine child outcomes, the matching estimator will be biased. An alternative estimation procedure that allows for selection on unobservables makes distributional assumptions on the unobservable characteristics in the child effect equation and the treatment equation. By specifying a bivariate normal distribution, and by having instrumental variables available that determine income treatment status, but not child outcomes, a *control function* can be added to the model for child outcomes that is equal to the conditional mean of the unobserved characteristics in the child outcomes model,

conditional on treatment status. By plugging in an estimate for the control function, this model can be estimated by OLS.

3.7 Difference in Differences approach to policy experiments

Where information is available over time (rather than just in a cross-section) a difference in differences (DiD) approach can be employed.⁶ This requires multiple observations per child (or household) over time both before and after the policy change, both for the group that was affected by the policy, and the group that was not. The DiD approach controls for differential household characteristics (family effects) - both observed and unobserved - that are constant over time, but cannot control for unobserved factors that are changing over time. This means that the success of this approach relies on the assumption that there are no such unobserved factors. For example, one has to assume that changes in macro-economic circumstances have the same impact on the two groups.

Let t_0 and t_1 denote the pre-and post policy change periods. Then, abstracting from other regressors an average outcome effect is estimated as

$$\left(\bar{Y}_{t_1} - \bar{Y}_{t_0}\right)^T - \left(\bar{Y}_{t_1} - \bar{Y}_{t_0}\right)^C$$

⁶ Instead of time, other natural groupings can also be used, like cohorts or siblings.

For this estimator we need to assume that the mean change in the non-treatment outcome is the same for treated and non-treated. In the model

$$Y_{it} = \beta_0 + \alpha D_i + u_{it},$$

let the error term be decomposed as

$$u_{it} = \eta_i + \lambda_t + \varepsilon_{it}$$

where η_i is an individual specific effect, constant over time, and λ_t is a common macro-economic effect, the same for all individuals. These effects are cancelled out by the differencing. If, however, the macro-economic effects differ between the treated and the controls the DiD estimator is biased. Also, if ε_{it} is correlated with D_i , DiD is biased.

Controlling for other factors, the DiD estimator is obtained by OLS on, for example,

$$Y_{it_1} - Y_{it_0} = \pi_0 + \alpha D_i + x'_{it_0} \pi + v_{it}.$$

Note that measurement error becomes important in the context of such estimators. A measurement error problem becomes more pronounced when variables are differenced. Thus bias due to measurement error is likely to be more severe here.

3.8 Sibling Differences

Another approach that has frequently been used to estimate the impact of parental income on child

outcomes is to try to difference out the effect of unobservable factors at the family-level by comparing outcomes across siblings (sometimes cousins). By comparing the outcomes of siblings who have been subject to different parental income at different ages, this approach allows the researcher to control for unobserved family-specific characteristics (such as parental ability) that are fixed over time. This approach is similar to the difference-in-differences approach we discussed earlier. Writing the outcome for child i in family f as:

$$Y_{if} = x_{if}\beta + \lambda M_{if} + \varepsilon_f + u_{if},$$

where we have decomposed the unobserved term into a family fixed effect ε_f (parental ability, affection, motivation etc) and a child specific effect, u_{if} , it is clear that if we have more than one child in a family, we can difference out the effect of fixed family characteristics, ε_f , by comparing siblings. In the context of trying to estimate the causal impact of income we have to be clear what we mean by “income in family f for child i ”. Unless they are twins, we would expect siblings to experience a different path of family income over their life – indeed for this approach to be successful the researcher needs to observe siblings sufficiently far apart in order for them to have experienced different financial circumstances.

As in all such models, the specification of parental income is an important feature of the model. If the researcher just controls for family income at a particular age, this would require us to believe that the outcome for the child at that age really only depended on that

period's income. More likely we would want to control for income differences at each age, or over an age range (e.g. 0-4 years, 5-10 years etc), but this information may not always be available from survey data – either we need to observe the family over a long period of time (at least from when their first child is born, and until subsequent children reach the ages at which we want to compare outcomes) or there need to be reliable retrospective income questions.

For the sibling difference approach to identify the causal impact of income (or any other variable) we must also be convinced that the family effect really is fixed, or that changes in it are not correlated with income, or we would be back to the same problem as with comparing outcomes between families. For example, unobserved parental motivation may affect both the parent's earnings and their parenting skills. If motivation really is fixed, then the sibling comparison can overcome the endogeneity problem. But if, for example, a parent becomes depressed, this is likely to both reduce their ability to cope with their job and with parenting, and sibling comparisons may no longer be effective for identifying the true impact of income. Similarly, if we think that u_{ij} contains a child specific idiosyncratic endowment (i.e. something innate to the child), then we need to consider whether differences in this endowment across siblings may be correlated with income differences (for example, severe learning difficulties in a child may make working more difficult for the parent thus reducing family income – although we might hope to have information on this sort of child characteristic), or whether the unobserved parenting inputs respond to sibling differences. Moreover, if there are child specific

differences then parents may attempt to compensate for them by substituting resources between the children.

Since this method is based on differencing it is likely that measurement error will be a further important source of bias.

4. Existing Evidence

There have been a large number of studies that have estimated models of child outcomes including income as an explanatory variable. We do not intend to provide a comprehensive review of the empirical literature here. (These can be found in Haveman and Wolfe, 1995, and Mayer, 1997). Instead we provide examples of a number of studies employing some of the methodologies discussed in the last chapter, and attempt to provide a summary of the main findings of these studies. Our reading of the main findings from the existing literature is that:

- The effect of current income on child outcomes is small;
- The effect of permanent income is much larger than the effect of current income, but usually decreases as more covariates are included;

- Income effects are small compared to the effects of race, gender, and many of the observable characteristics of the parents – it would take large financial transfers to overcome the disadvantage associated with certain characteristics;
- The income effect is generally found to be non-linear, with larger effects at lower levels of income – having said this, few studies allow for such non-linearity.
- Comparing the effects of different policy interventions suggest that income at different ages matter differentially – with outcomes being more sensitive to variations in income at a young age relative to at older ages.
- Studies employing methodologies to take into account the unobservable characteristics of families and children in general find much smaller effects of parental income on child outcomes than those that do not.

In the sections below, we review a number of studies that apply some of the econometric methods discussed in this report to measure the impact of parental income on child outcomes.

4.1 Ordinary Least Squares, and other methods that do not identify causal effects when income is endogenous

There is a relatively large body of literature using conventional estimation methods such as OLS to estimate the impact of parental income on child

outcomes. In addition most papers that use the alternative methods discussed in this report start by presenting results using OLS, before contrasting these with their results which attempt to identify causal effects.

Mayer (1997) is a distinctive study that is notable for its extensiveness and with the variety of ways in which she carefully attempts to uncover the causal impact of income. The conclusion of her detailed study mostly using the US PSID is that while conventional OLS estimates of the effect of long-run family income imply that some effects of income are sizable⁷, such sizeable effects are more or less eliminated once methods better suited to measure the casual impact of income are adopted. Her conclusion is that there is no persuasive evidence that additions to income produce large improvements in child outcomes.

One careful example of the application of simple least squares methods, albeit with a rich dataset, is Korenman, Miller, and Sjaastad (1995), one of several papers that uses the US NLSY data. Their estimates indicate that, for a range of behavioural and academic outcomes, a \$1,000 increase in current income (in 1993 dollars) is associated with fairly small effects – though larger than the effects found by Mayer once she moves beyond OLS to more sophisticated methods. The largest impact they found, on reading ability, was 2.35% of a

⁷ Mayer's conventional estimates suggest, for example, that doubling annual family income from \$15,000 (approximately the poverty line) to \$30,000 reduces out-of-wedlock childbearing by 18 percentage points and cuts the high school dropout rate by almost 13 percentage points. It boosts young men's earnings by \$4,400 a year - more than a quarter.

standard deviation (s.d.)⁸. (When some of the control variables were removed, the effects get larger.) The income effects found in this paper were non-linear. For example, the average effect across five different outcomes was 2.16% of an s.d. of the dependent variables evaluated at an income of half the poverty line, but averaged only 0.7% of an s.d. when income was between 1.85 and three times the poverty line.

Results using conventional methods from three other data sources also yield findings similar to those from the NLSY. Duncan *et al.* (1994) use data on premature children from the Infant Health and Development Program, and estimate, again using simple regression methods but with limited family background controls, that a \$1,000 increase in (four-year average) family income would increase IQ recorded at age 5 by 1.5% of a s.d., and reduce a behaviour-problems index by approximately 0.7% of a s.d.

Hanushek (1992) uses data from the Gary Income Maintenance Experiment on a sample of low-income black families to estimate, again using simple regression methods, the determinants of the change in achievement between the second and sixth grades as a function of income (averaged over five years). The results, controlling for the number of children in the family, teacher fixed effects, lagged achievement, and a variety of other behaviours and outcomes, suggests that an \$1,000 increase in income would result in an increase of 1.8% of an s.d. of the educational achievement change.

⁸ Assessing the impact of a given income transfer on the number of standard deviations of an outcome gives us a useful metric to compare coefficients across different outcomes and different studies.

Wolfe (1981) uses data from the Collaborative Perinatal Project, a mostly poor black sample from Philadelphia. She finds that, in a regression model that includes only income and parental education, a \$1,000 change in income at age 7 would change age-7 IQ by about 2.0% of a s.d. but that adding additional control variables reduces this effect to about 1.0% of an s.d.

Evidence from the UK again using conventional methods suggests perhaps a greater role for parental income in determining child outcomes. Hobcraft and Kiernan (2001), examine the relationship between childhood poverty and later outcomes, such as teen motherhood, social exclusion, health, and emotional well-being (malaise and life satisfaction) in the UK National Child Development Study. They find large effects of being “clearly poor” in childhood – for example doubling the odds of living in social housing and raising the odds of an extra-marital birth of around 50%. However, the analysis makes no attempt to investigate the causal impact. Another example of detailed research using NCDS is Gregg and Machin (1998) who use regression methods to find relatively large effects of parental financial distress on the education and labour market outcomes of young adults, but again they do not attempt to identify causal effects.

The empirical literature does seem to suggest that income has different effects at different ages. For example differences in test scores across parental income are apparent at early ages in Carneiro and Heckman (2002). They show significant differences in mathematics test scores across parental income quartiles when the child is age 6, and they show that these test score differences widen over the following six years.

Even controlling for a large set of background factors such as parental education and mother's test scores did not eliminate the effect of income on maths test scores. The effects are also found to be important for the long run. Carneiro and Heckman go on to show that the maths test score at age 12 and long-run average family income over the entire childhood have significant effects on the probability of attending college. Indeed, they argue that it is these long run factors that affect college enrolment - income immediately prior to college (ages 16-18) has no effect.

Keane and Wolpin (2001) also fail to find an effect of borrowing constraints amongst young adults on college enrolment. However they do find that students from low current income families are more likely to work while in college in response to the credit rationing. So relaxing such constraints would mainly affect their market work.

Another recent paper using conventional methods to examine the relationship between parental income and child health is Case *et al* (2002). This paper suggests that US parental income, averaged over a number of years, is positively related to child health and the relationship becomes stronger as the child ages. Why this is the case is unclear. The authors suggest that higher income parents are better able to cope with the incidence of poor health and so their children recover better. This results in better long term health and less lost schooling. The authors conclude that providing poor parents with more income may not do much to change the ability of parents to make these desirable investments.

4.2 Instrumental variables

An example of using instrumental variables to examine the impact of parental income on child outcomes is Shea (2000). He uses data from the US Panel Study of Income Dynamics (PSID). He argues that union status, industry, and involuntary job loss gives rise to income variation that is not correlated with unobserved hereditary ability that is likely to affect both parental earnings and child earnings. These variables are therefore used as instruments for income in the child outcome regression.

In general, Shea finds significant, positive effects of parents' income on children's outcomes in the OLS regressions, but no significant effects when moving to the IV regressions. Indeed, the point estimates of the coefficients on parents' income are often negative, but they tend to be very imprecisely estimated.

However one difficulty with this technique is that it is often hard to think of plausible instruments. Shea himself concedes that union and industry premia may partly be due to ability. If this is the case then union status would be an invalid instrument, and the results from instrumenting in this way would be biased. There is also some evidence in Shea's paper that the instruments may be weak⁹. The problem of weak instruments was discussed in section 3.1.

⁹ The (partial) R-squared for the first-stage regression - this is a measure of how much variation in the endogenous variable is explained by the instruments once the effect of the other exogenous variables have been allowed for - is rarely above 10% suggesting that the instruments are not very strong.

Mayer (1997) uses a number of methods exploiting potentially exogenous income variation to understand the impact of income on child outcomes. In one chapter of her book she effectively uses growth in parental income after an outcome has occurred (“income after the outcome”), as an instrument for initial parental income. The idea is that if future income growth is unanticipated, or is anticipated but that peoples’ responses to changes in future income are random with respect to unobserved characteristics that determine children’s outcomes, then actual growth in parental income can be used as a source of exogenous variation.

Using this approach, Mayer (1997) finds that “all else being equal, high parental income hardly affects children’s behaviour problems or reading scores” but that the effect on maths scores is “slightly larger than the conventional estimate”. However, none of the estimated coefficients are significantly different from zero.

However, when the approach is viewed in terms of instrumental variables, the drawbacks become apparent. For the difference in income between two time periods, to be a valid instrument for income in period 1, it is necessary that this income difference is independent of the unobserved characteristics that jointly determine income levels and children’s outcomes (e.g. motivation to succeed). This seems very doubtful. Furthermore, simple life-cycle models suggest that if income changes are anticipated, then these changes are almost fully incorporated into their decision making process well in advance of them actually occurring. This suggests that future income is in itself an endogenous factor that may need to be instrumented.

4.3 Experimental evidence

There have been relatively few examples of randomised income experiments that allow us to assess the impact of income transfers to parents on child outcomes. One example is the Minnesota Family Investment Program (MFIP), used in Morris and Gennetian (2002) to investigate the impact of parental employment and income on children's school achievement and behavioural problems.

The MFIP was a randomly assigned trial, in which single parents in seven Minnesota counties who were long-term recipients of welfare were randomly assigned to one of 3 groups¹⁰. Group 1 - MFIP Incentives Only - were allowed to retain their standard benefit payments as their earnings increased. This was intended to encourage parents to commence work and seek to improve their labour market outcomes. Group 2 - Full MFIP - received these same incentives, and in addition were required to participate in employment and training services. This is in contrast to the standard benefits received by the controls, Group 3, where AFDC welfare payments were reduced dollar for dollar with earnings.

The effects of MFIP were estimated by following the three groups over time and comparing their employment, welfare receipt, and other outcomes, including child outcomes. The authors identify two possible causal mechanisms through which the program could have an impact on children's outcomes. First,

¹⁰ Between April 1994 and March 1996, more than 14,000 families were randomly assigned.

there is one channel through the effect of increased incomes of families in the treated groups, and second, there is another channel through the impact of increased employment among the same families. The benefits of extra income may be offset by the potential decrease in the quality of parenting that comes as a result of longer hours being worked. However, the design of the experiment, which incorporates two levels of random assignment enables the use of indicators for assignment to each of Group (1) and Group (2) as two instruments for the two potentially endogenous variables in the children's outcomes equation – income and employment.

The findings show slightly significant positive effects of income on indicators of children's school engagement and positive behaviour, as reported by the mother 3 years after random assignment. However these results only refer to the effect of income in the year after random assignment. When income is measured over the 3 years after random assignment, the low precision of the estimates mean that the effect on these two outcomes is not significant.

4.4 Examples of Natural Experiments

Relatively few papers have exploited natural experiments to identify the effects of changes in parental income on children's outcomes - certainly none have done so in a UK context¹¹

¹¹ The evaluation of the Education Maintenance Allowance looks at the impact of an education subsidy – i.e. money tied to educational attendance – on post-compulsory school attendance. It is therefore not appropriate for estimating the impact of income more generally.

Duflo (2000) is an example of exploiting a natural experiment to estimate the effects of income on children's outcomes. Here, the experiment is the increase in the generosity of the old age pension (OAP) to black South Africans in the early 1990's. The OAP is a universal (given pensionable age) means-tested benefit. This represented an exogenous, permanent increase in income for the eligible group, many of whom lived in families with children present. Duflo's aim is to test whether this increased income impacted on children's nutrition in households with a pension recipient, as measured by weight for height/age for children up to five years of age. Simply comparing outcomes across households with and without pension recipients could be misleading, since households with no pension recipient tend to be better off to start with than those with a recipient. A DiD-type estimator is possible here, despite the fact that Duflo only has a single cross-section of data from 1993. The reason is that this data contains information on children aged 0-5 years, the oldest of whom will only have been affected by the pension increase for part of their lives. Since current weight for height/age depends on past nutritional inputs, we expect a lower impact of the program on the outcome variable for older children than for the younger children exposed to the program all their lives. The assumption necessary to identify the income effect is that, in the absence of the pension reform, the differences in child outcomes between eligible (treatment) and non-eligible (control) families for the "young" children would have been the same as it is for the "old" children¹².

¹² Hence the impact of the reform will probably be under-estimated, unless we think there was no effect at all on the older children.

Duflo was careful to distinguish between the effect of income to grandfathers from income to grandmothers and between granddaughters and grandsons. She found no significant effects of income to grandfathers on the outcomes of either grandsons or granddaughters. But she did find quite significant effects of income to grandmothers on the outcomes for granddaughters but no significant effects for grandsons. However, it was unclear why such differences occur.

A number of other papers have attempted to use variation in benefit payments across both place and time to estimate the impact of parental income on child outcomes. For example, Mayer (1997) applies the strategy of exploiting cross state and cross time variation in Aid To Families with Dependant Children (AFDC). AFDC is a means-tested benefit that almost exclusively serves single-parent families. There are large variations in the level of the benefit across states¹³,

However, whether the level of AFDC is truly exogenous with respect to children's outcomes is questionable, as there are numerous other factors (both observed and unobserved) that differ across states, and that are correlated with both the children's outcomes and AFDC levels. To control for these differences, Mayer uses the fact that outcomes for children in two-parent families could not be affected by the level of AFDC, but will be driven by the same state-level characteristics that affect the outcomes for children in one-parent families. The identifying assumption then, is that interstate differences in the gap between outcomes for children in

¹³ Ranging from \$680 for a family of three in Connecticut to \$120 for a similar family in Mississippi at the time of Mayer's study.

one- and two-parent families is only affected by interstate differences in the level of AFDC. This is equivalent to assuming that state-level factors, other than the level of AFDC benefits, affect outcomes for children in one- and two-parent families in the same way.

Mayer estimates this model for the following seven outcomes, using data from the Panel Survey of Income Dynamics (PSID): years of education, wages, earnings, and probabilities of teenage childbearing, dropping out of school, male idleness and single motherhood. The only outcomes for which the effects are statistically significant are dropping out of school and single motherhood.

Another recent paper to exploit US state-time variation in cash welfare payments is Paxson and Waldfogel, 1999, who look at the relationship between child abuse and neglect and welfare payments. They find that reductions in welfare benefit levels are associated with small increases in child maltreatment.

4.5 Sibling differences and other fixed effects models

One interesting paper, which uses a number of different methods, including sibling differences and child fixed effects models to estimate the impact of mother's welfare receipt on several developmental outcomes, is Levine and Zimmerman (2000). Using data from the NLSY, the paper finds that simple correlations suggest a strong negative relationship between maternal welfare receipt and children's outcomes. However once they adopt more sophisticated estimation strategies, designed

to identify whether this correlation can be attributed to the mother's welfare receipt directly or to other characteristics of mothers who receive welfare, regardless of whether or not those characteristics are observable to the researcher, the authors conclude that they can find "little evidence" of any causal link between maternal welfare receipt and children's developmental outcomes.

Duncan et al (1998) studies the effect of family income on completed years of schooling with the US PSID data. In addition to looking at conventional comparisons across families, they also compare siblings in the same family. They find positive effects of family income on completed schooling, particularly when the child was young.

This finding is supported in Ermisch, Francesconi and Pevalin (2002), for the UK. They use a measure of poverty based on 'parental joblessness' and use this as a basis for comparisons across siblings. They find that early poverty significantly reduces the probability of achieving at least one 'A-level' pass – a qualification used to determine university entrance.

Garces, Thomas and Currie (2002) is an example of the use of a family fixed effect model. They are not specifically concerned with the effect of income but with the effect of "Head Start" attendance on outcomes of children later in life. Head Start is a public pre-school programme in the US aimed at disadvantaged children. The study uses a 1995 supplement to the Panel Survey of Income Dynamics, which asked retrospective questions about participation in Head Start and other pre-schools to respondents aged 18 to 30. Four

subsequent outcomes were considered: completing high school; attending college; earnings, and being charged with a crime. A simple cross-section regression (controlling for attendance at another pre-school and a limited set of factors such as race and year of birth) shows that Head Start attendance tends to be statistically significantly associated with worse outcomes, which is hardly surprising since the program is targeted towards disadvantaged children. Adding further covariates associated with family background such as parental education, family income when the child was aged four, birth weight and so on tended to reverse the finding that attending Head Start is associated with worse outcomes, although the effect generally became insignificant. Similar results are found using a family fixed effects model and the coefficients were generally found to be even less precise.

The family fixed effect model compares siblings one of whom has received Head Start and the other has not. This begs the question, why parents would send one child to Head Start and not the other. It could simply be that family circumstances change so that they are no longer eligible for Head Start and this should be controlled for to some extent by the inclusion of family income and attendance at other pre-schools. Of more concern is when both siblings are eligible but only one attends. For example, if this selective use of Head Start is a compensating response to idiosyncratic child endowments then the results could be biased.

4.6 The effect of other inputs on child outcomes: parental employment and neighbourhood effects

Given the close relationship between parental work and income, it is very important to know if the effect of one offsets the other in terms of childhood outcomes—especially if policy is directed towards reducing disadvantage through promoting work.

Much early research seemed to point to negligible effects of parental employment on children. However Ruhm (2000) investigates the relationship between parental employment and child cognitive development in data from the National Longitudinal Survey of Youth. Maternal employment during the first three years of the child's life is predicted to have a small negative effect on the verbal ability of 3 and 4 year olds and a substantial detrimental impact on the reading and maths achievement of 5 and 6 year olds. The results further suggested that paternal employment effects were similar.

Ermisch and Francesconi (2002) also study this issue and use a sample of siblings born in the 1970s. They find a significant negative effect of mother's full-time employment (smaller for more-educated mothers) when the child was aged 0-5 on the probability that they received at least one A-level but no effect of mother's part-time employment or father's employment. The smaller effect for more educated mothers might reflect their better capacity to manage the impact on the child, partly through market childcare.

Another important factor when considering the impact of income on child outcomes is neighbourhood quality. Whether one wants to control for this at all is debatable. If we are interested in the effects of providing greater resources then it may be inappropriate to control for

neighbour (or school) quality if some of those resources were to be spent by parents on improving the quality of their neighbourhood. On the other hand poor households may face very limited choices and indivisibilities may suggest that additional income transfers may not change these fixed circumstances. In which case it may be appropriate to control for such differences.

An example of a paper examining the role of neighbourhood effects is Brookes-Gunn *et al* (1998). This paper considers whether, apart from parental background, neighbourhoods also matter for child outcomes such as IQ, teenage motherhood, and dropping-out from school. The neighbourhood measure used is average income within neighbourhoods. They use OLS and do not correct for neighbourhood sorting. The authors suggest that the presence of higher-income households has a positive influence on an index of child outcomes but there is no corresponding negative impact of lower-income households. This paper, along with almost all other investigations, find that the role of the family is important and that when family effects are included neighbourhood effects become somewhat less important. However, their results are vulnerable to the criticism that there is endogenous population sorting. Families who live in high income neighbourhoods may do so because of unobserved characteristics correlated with parental success, so that the neighbourhood measures are endogenous. In fact, they argue that sorting is unimportant and note that the bias created by any sorting could be up or down - although most researchers have taken the position that sorting is most likely to bias estimates of neighbourhood effects upward.

5. Conclusion

Policymakers and many researchers have been prone to take it as given that more resources improve welfare. This review has been concerned with what seems like a simple question “How much does it improve welfare”? We have been generally pessimistic about attempts to answer this. Our review has outlined just how difficult it is to address even this simple question and has presented the best available evidence.

This report has set out some commonly used techniques for estimating the impact of parental income on child outcomes. Since parental income is likely to be endogenous to child outcomes because of unobservable factors, one of two basic approaches needs to be adopted. The first approach is either to create - by means of an experiment - or to search for *exogenous* variations in income. Such variation in income will not be correlated with parental (unobservable) characteristics. The second approach is to difference out the effect of unobservables, generally on the assumption that they are fixed over time, by comparing outcomes for the same households, or similar groups of households, over time.

A reasonable reading of the empirical literature – mostly from the US - suggests that the effects of permanent family income on child development outcomes are generally probably too small to make income transfers to low-income households a sensible approach to generating large changes in outcomes for low-income children. While no one method is completely convincing, on balance the alternative estimates strongly suggest that conventional OLS estimates overstate the impact of income, often by a substantial

margin. This is particularly true for those outcome variables where the conventional estimates imply that the effect of family income is very large.

Of course the use of both time and money resources is usually controlled by parents and if parents are committed to the well-being of their children and if they are competent at allocating resources then we would expect that additional resources, say from welfare transfer programmes, would imply better outcomes for children. Equally, if parents are not competent in terms of productivity within the household, or simply do not place weight on the well-being of their children, then additional resources might be expected to have only limited effects on children's current welfare and their later success as adults. While we have no way of knowing about how parental domestic productivity or their attitudes towards their children vary, it would not be very surprising if there turned out to be a large variance across households in how effective household resources are used in improving outcomes for children.

Family background and other family and child characteristics usually have large effects on child development (and later adult outcomes). These findings suggest that the main inputs to the production of child development are fixed ones that are not greatly affected by changes in income. For example, later economic success is much more closely related to family background than it is to family income when young.

Some family background factors, such as parental education, have moderate effects on child development but larger effects on later adult outcomes which suggests that the main mechanisms through which they

affect adult outcomes are not particularly well captured by the child-development outcomes. In other words, the mechanism through which parental education, for example, might affect the later adult outcomes of their children is through promoting better choices such as later school leaving or through encouraging good social skills¹⁴.

The literature further suggests that many government income-support programs have little direct impact on child development, and that the public provision of health and education services children may be the most effective means of improving child development. Thus, while it is clear that there are sizeable differences between the outcomes experienced by children across the range of parental incomes, the evidence mostly implies that income does not *cause* these differences. The implication is that policies that increase the income of poor parents will result in a small increase in child outcomes compared to the observed differences in child outcomes across the range of parental incomes. Income transfer programmes are not a quick fix for poor child outcomes.

However, while the literature has shown few sizeable effects on any one outcome there is a general finding of small effects over a range of outcomes. No studies have yet attempted to quantify the combined effect across all studies – not least because we have no way of weighting

¹⁴ Even here, there is recent evidence that casts doubt on earlier studies. For example Berhman et al, 2002, use a large sample of Minnesota twins who are mothers and compare their differences in schooling with that of their children. They find a negative correlation between mothers' and childrens' schooling in some specifications and no significant positive effects in any specification.

them together until we can put a financial value on the worth of a change in any particular outcome. In other words, a comprehensive evaluation of the effect of parental income requires a fully-fledged cost benefit analysis that quantifies all effects in monetary terms.

References

Adam, S. and Brewer, M. (2004), *Supporting Families. The Financial Costs and Benefits of children since 1975*. Published for the Joseph Rowntree Foundation by the The Policy Press

Becker, G. and N. Tomes, (1986), "Human capital and the rise and fall of families", *Journal of Labor Economics*, 4(3), S1-S39.

Black, S.E., P.J. Devereux and K.G. Salvanes, (2003), "Why the apple doesn't fall far: understanding intergenerational transmission of human capital", *National Bureau of Economic Research Working Paper* 10066.

Blundell, R. and M. Costa Dias, (2000), Evaluation Methods for Non-Experimental Data, *Fiscal Studies* 21, 427-468.

Bound, J., D.A. Jaeger and R.M. Baker, (1995), "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association* 90, 443-450.

Brooks-Gunn, J, G.J. Duncan, P. Klebanov, and N. Sealand. 1993. "Do Neighborhoods Influence Child and Adolescent Development?", *American Journal of Sociology*, 99(2), 353-395.

Cameron, S. and J.J. Heckman, (2001), "The Dynamics of Educational Attainment for Black,

Hispanic, and White Males”, *Journal of Political Economy*, 109(3), 455-499.

Carneiro, P. and J.J. Heckman, (2002), “The evidence on credit constraints in post-secondary schooling”, *Economic Journal*, 112(482), 705-734.

Case, A., D. Lubotsky and C. Paxson (2002), “Economic status and health in childhood: the origins of the gradient”, *American Economic Review*, 92, 1308-1334.

Currie, J. and D. Thomas, (1995), “Does Head Start Make a Difference?” *American Economic Review*, 85(3), 341-364.

Duncan, G.J., J. Brooks-Gunn and P.K. Klebanov, (1994), “Economic development and early childhood development”, *Child Development*, 65, 296-318.

Duncan, G.J., W. Young, J. Brooks-Gunn and J. Smith (1998), “How much does childhood poverty affect the life chances of children?” *American Sociological Review*, 63, 406-423.

Duflo, E., (2003), “Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa”, *World Bank Economic Review*, 17(1), 1-25.

Ermisch and Francesconi (2002)

Ermisch, J., M. Francesconi and D. Pevalin (2002), “Childhood parental behaviour and young people’s

outcomes”, *ISER Working Paper 2002-12*, University of Essex.

Garces, E., D. Thomas, and J. Currie (2002), “Longer-Term Effects of Head Start”, *American Economic Review*, 92, 999-1012.

Gregg, P. and S. Machin, (1998), “Childhood disadvantage and success or failure in the youth labour market” *Centre for Economic Performance Discussion Paper 397*.

Hanushek, E.A., (1992), “The Tradeoff between Child Quantity and Quality”, *Journal of Political Economy*, 100(1), 84-117.

Haveman, R. and B. Wolfe, (1995), “The determinants of children’s attainments: A review of methods and findings”, *Journal of Economic Literature*, 23, 1829-1878.

HM-Treasury, (2001), “Tackling child poverty: giving every child the best possible start in life”. Pre-Budget Report Document, December 2001.

Imbens, G.W. and J.D. Angrist, (1994), “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62(2):467-75.

Keane, M. and K. Wolpin (1997), “The career decisions of young men”, *Journal of Political Economy* 105:473-522.

Korenman, S., J.E. Miller, and J.E. Sjaastad, (1995), “Long-Term Poverty and Child Development in the

United States: Results from the NLSY”, *Child and Youth Service Review*, 17(12), 127-155.

Mayer, S., (1997), “What Money Can't Buy: Family Income and Children's Life Chances” (Cambridge, MA: Harvard University Press).

Paxson, C. and J. Waldfogel, (1999), “Work, Welfare, and Child Maltreatment”, *National Bureau of Economic Research Working Paper* No. w7343, Issued in September 1999.

Ruhm, C.J. (2000), “Parental employment and child cognitive development,” *National Bureau of Economic Research Working Paper* 7666.

Shea, J., (2000), “Does parents' money matter?”, *Journal of Public Economics*, 77(2), 155-184.

Wolfe, J.R., (1981), “The Impact of Family resources on Childhood IQ”, *Journal of Human Resources*, 17(2), 213-235.