View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Faculty Research Working Papers Series

On the Measurement of Poverty Dynamics

Daniel Hojman
John F. Kennedy School of Government - Harvard University

Felipe Kast
Universidad Catolica de Chile

November 2009
RWP09-035

The views expressed in the HKS Faculty Research Working Paper Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

On the Measurement of Poverty Dynamics*

Daniel A. Hojman
Harvard University

Felipe Kast
Universidad Catolica de Chile

This Version: October, 2009; Preliminary: August, 2007

Abstract

This paper introduces a family of multi-period poverty measures derived from commonly used static poverty measures. Our measures trade-off poverty levels and changes (gains and losses) over time, and are consistent with loss aversion. We characterize the partial ranking over income dynamics induced by these measures and use it in two empirical applications with longitudinal household level data. Comparing two decades of income dynamics in the United States we find that the income dynamics of the 1990s -post Welfare reform- dominates the income dynamics of the 1980s -pre Welfare reform. Next, we compare the contemporary income dynamics of three industrialized countries and conclude that United Kingdom dominates Germany and United States, and Germany dominates the United States if poverty stocks are given more importance than poverty flows. The differences between our ranking and those obtained using other welfare criteria such as social mobility suggest that our measures capture critical information about the evolution of poverty.

JEL Classification: D03, D63, I32
Keywords: poverty measures, poverty dynamics, social mobility, loss aversion

*We thank Alberto Abadie, Filipe Campante, David Ellwood, James E. Foster, Ed Glaeser, Peter Gottschalk, Sandy Jencks, Asim Khwaja, Dan Levy, Jeffrey Liebman, Erzo Luttmer, Nolan Miller, Dina Pomeranz, Monica Singhal and Richard Zeckhauser, and seminar participants for comments and suggestions. Michelle Favre and Sungchun No provided excellent research assistance. Hojman is grateful to the Taubman Center for State and Local Government for financial support. Kast is grateful to the Harvard Multidisciplinary Program in Inequality & Social Policy for financial support.

1 Introduction

In modern societies there is substantial mobility in and out of poverty. Over the last two decades the dynamics of poverty has been the subject extensive empirical research.[1] This work has changed our understanding of poverty by quantifying its persistence, and identifying the factors more likely to determine an individual's ability to escape poverty and the events likely to trigger poverty over the life cycle. The findings have strongly influenced the reform of poverty alleviation programs in the United States, Great Britain, and other industrialized countries in recent years.[2]

Even though the dynamic dimension of poverty has inspired of a body of empirical research and has influenced policy design, the theory of poverty measurement has lagged behind.[3] This paper introduces a family of multi-period poverty measures derived from commonly used static poverty measures. We use these measures to rank *income processes* focusing on the dynamics of poverty. The framework is used in two applications. First, we compare poverty dynamics across two decades in the United States -the eighties and the nineties. Second, we rank poverty dynamics across three industrialized countries -Germany, Great Britain and the United States. The method delivers a significantly different ranking than the ones that rely either on static poverty changes or social mobility measures.

Our framework builds on recent work on *multidimensional poverty* measurement and some of the most robust findings in behavioral economics. We assume that the well-being experienced by the individual over time is determined by the stream of a "welfare attribute" over time. In keeping with the poverty literature, this welfare attribute is referred to as *income* and the stream as an *income path* or *trajectory*.[4] A society is described by the profile

[1] A seminal paper is Bane and Ellwood [1986] who estimate the persistence of poverty spells in the U.S.

[2] A central aspect of the Welfare Reform in the United States introduced during the nineties has been to make welfare transfers conditional on the beneficiary's participation in the labor market or work-related activities such as training. An underlying principle was to promote self-sufficiency over time. See Blank [2002] for an analysis of the U.S. reforms and Hills [2004] for an overview of the Britsh reforms.

[3] Thorbecke [2004] argues that most of the unresolved issues in poverty analysis are related to the dynamics of poverty. See also Kanbur [2005] and a recent collection of essays in Addison, Hulme, and Kanbur [2009].

[4] The framework allows for this attribute to be a vector capturing a wide range of dimensions of well-being including the consumption of different goods and services, en-

of income paths for each member of society referred as an *income dynamics.* A multi-period poverty measure is an index that assigns a number to such profile. We consider multi-period measures that are consistent with an underlying static derivation scale. A static deprivation measure assigns a measure of deprivation to each income level. Thus, each individual income path can be associated to a *deprivation path*, the stream of deprivation levels associated to each period. The profile of all deprivation paths in society defines a *poverty dynamics.*

The multi-period measures proposed in this paper satisfy the core axioms that characterize static poverty measures. In addition, we introduce axioms that bear on the dynamic nature of our task. In particular, the paper offers an axiomatic foundation for measures that allow individual well-being to depend on both the levels of a welfare attribute and also its changes over time. The latter builds on the literature on reference-dependence that stresses the importance of changes as carriers of utility, as in Kanheman and Tversky's [1979] classic work on prospect theory. Our three main axioms are *monotonicity, stock-flow separability,* and *loss aversion.* Monotonicity refers to the fact that lower levels of the welfare attribute are reflected in higher multi-period deprivation. The stock-flow separability axiom implies that measures can be expressed as a function of levels and changes of the welfare attribute. The loss aversion axiom captures the idea that, given income streams with the same levels of deprivation but in a different sequence, an individual is better off with an increasing sequence of outcomes than a decreasing one. For illustration, suppose that at each period the deprivation of an individual is summarized by an indicator of whether or not the individual is poor given her income level.[5] Thus, over two periods of time, there are four possible deprivation trajectories: An individual can be poor in both periods -*always poor*, non-poor in both periods -*never poor*, start poor and end non-poor -*poverty outflow*, or start non-poor and end poor -*poverty inflow.* Monotonicity implies that the *always poor* and the *never poor* paths are respectively the worse and best paths. The other two paths involve a change in poverty status over time and, in the absence of further restrictions, the relative ranking of these paths is unclear. The *loss aversion axiom*

dowments, and measures of psychological and physical health, among others. It should be clear however that our primary source of multidimensionality is the consideration of attributes -possibly a single one- over multiple periods of time.

[5] This example assumes that the underlying static deprivation scale is the poverty "headcount".

postulates that paths associated to poverty inflows have lower experienced well-being than those associated to poverty outflows.[6]

Theorem 1 in section 4 provides a complete characterization of the individual multi-period measures that satisfy our axioms. At the aggregate level we show that the multi-period poverty measures defined by our axioms can be decomposed into two terms (Lemma 3). The first term is population average of an increasing function of the levels of the welfare attribute. The second term is a population average of a function that evaluates gains and losses between consecutive periods. We use this representation theorem to provide a complete characterization of the partial ranking induced by these measures on the space of poverty dynamics. The characterization obtains from solving an optimal control problem and the partial ranking we derive is determined by a set of stochastic dominance conditions. For example, If the underlying static deprivation scale used is the headcount or poverty indicator, society A dominates B if three conditions are satisfied. First, there are more individuals that are never poor in A than in B. Second, the level of final poverty -the sum of those who are always poor and those who enter poverty- is lower in A and B. Third, there are less individuals that are always poor in A than in B. Hence, the welfare criterion implied by our measures is determined both by the "stocks" and "flows" of poverty over time.

Two empirical applications using longitudinal household level data are developed. We first compare two decades of income dynamics in the United States and find that income dynamics of 1990s dominates the income dynamics of the 1980s. Next we compare the contemporary income dynamics of three industrialized countries and conclude that Great Britain dominates both the United States and Germany. It is not possible to rank the United States and Germany for all the measures consistent with our axioms. Indeed, for measures that give sufficiently high weight to poverty inflows and outflows relative to poverty stocks, the United States ranks better. Conversely, if the measure gives lower relative importance to poverty creation and destruction than poverty stocks, Germany is favored in the comparison. As discussed in detail in the sequel, the applications illustrate that the ranking produced by our method can be quite different than those based on social mobility.

[6] At the individual level the axiom consistent with the findings of behavioral economics, including the preference for improving outcome sequences with commensurable aggregate outcomes, recent evidence on the evolution subjective well-being showing that it is easier to adapt to a positive income shock than a negative shock, and, of course, loss aversion. We summarize this evidence in section 2.

Our paper is closely related to the poverty measurement and the social mobility literatures. For illustration, consider a two-period society. The domain of our multi-period poverty measures is the set of bivariate distributions of two-period income paths. Given any distribution f over income paths the income distribution in period $t \in \{1, 2\}$, f_t, is just the marginal distribution of f for period t. Static poverty measures rank income distributions and cannot be applied to rank income processes but can be used to measure the change in poverty from distribution f_1 to f_2. The same poverty change will be generated by any combination of inflows and outflows with a constant difference.

On the other hand, social mobility measures focus on the transition properties associated to f. For any income path (y_1, y_2), we can write $f(y_1, y_2) = f_1(y_1)W(y_1, y_2)$, where f_1 is the marginal distribution of income in period 1 defined above and $W(y_1, y_2)$ is the conditional probability of a transition from income y_1 in period 1 to income y_2 in period 2. In most of the social mobility literature the focus is on describing the properties of the transition matrix W. In comparing two different income dynamics the tendency is either to ignore the base rate f_1 or, as in Atkinson's seminal welfare-based approach to social mobility (Atkinson [1983]), to assume that the relevant comparison is across societies with the same marginal distribution but possibly different transition matrices.

This suggests several differences between the welfare criterion implied by our approach and the social mobility literature, as confirmed by our applications. First, multi-period poverty measures depend not only on income transitions described by W but also on the stock of people who are poor f_1. Second, our measures focus on the mobility in and out of poverty rather than mobility across the entire distribution of income. Third, a consequence of monotonicity is that poverty outflows increase welfare but poverty inflows decrease welfare. In contrast, the welfare-based approach to social mobility that followed Atkinson's seminal work is guided by the principle of equalizing opportunities,[7] which leads to a favorable view of societies characterized by high "circulation", i.e., those with large numbers of individuals both rising

[7] One interpretation of this principle is the notion of "origin reversal" which captures the idea that an income process is more desireable to the extent that an initial position in the income distribution is easily reversed (Dandardoni [1993]). Another interpretation rest on the notion of "origin independence" which captures the idea that an income process is more desireable to the extent that future well-being is independent of an individual's initial income. See Gottshalk and Spolaore [2001] for a discussion.

and falling in the income distribution. We argue that, by design, our measures are better suited to reflect the evolution of well-being of the poor. This is not a critique of social mobility measures but it emphasizes that our measures are guided by a different normative benchmark and capture different information, one we believe might be relevant for policy and research.

Finally, our paper also contributes in two active areas of research in welfare economics. The first of these is the recent literature of multidimensional poverty and poverty over time.[8] Foster [2007] proposes a definition of poverty over time. In contrast, as discussed shortly, we side-step the issue of identifying the "poor over time". Instead we axiomatize a family of multi-period poverty measures consistent with any definition of poverty over time. Our focus is on characterizing the ordering induced by this family on income dynamics. Atkinson and Bourbignon [1982] and more recently Bourbignon and Chakravarty [2002] and Duclos et al. [2006] study the orderings associated to multidimensional poverty measures. These papers provide a characterization for measures that satisfy properties other than the ones considered here and, more importantly, they do not focus on the dynamic dimension of poverty. As argued in section 2, the dynamic dimension may require specific normative guidelines as explored in this paper.

At the same time our work contributes to a growing literature that informs applied welfare analysis with the findings of behavioral economics. Other examples in this vein include Kanheman and Sugden [2005], Kanheman and Krueger [2006], and Chetty [2009a,2009b]. To the best of our knowledge, this is the first paper to provide an axiomatic poverty framework with axioms founded on evidence from the field of psychology and economics.[9]

The rest of the paper is organized as follows. In Section 2 we discuss the empirical evidence on well-being over time, other normative aspects associated to the dynamic nature of our framework, and overview the main properties of our measures and the ranking they induce with an example. Section 3 introduces multi-period poverty measures. In Section 4 we introduce the main axioms and provide a representation theorem of the measures that satisfy them. Section 5 characterizes the partial ordering induced by

[8] An important paper in general multidimensional poverty measures is Bourbignon and Chakravati [2003].

[9] While the nature of this paper is normative, as discussed in section 4, our measures are related to the reference-dependent preferences over consumption streams introduced in Gilboa [1989] and Shalev [1997]. See also Bowman, Minehart and Rabin [1999], Koszegi and Rabin [2006], and Rozen [2009].

these measures. Our applications are in section 6 and 7.

2 Well-being over Time: A Normative Baseline

In this section we provide some foundations for the assumptions that underlie our social welfare criterion. We focus on those that are specific to the dynamic nature of our problem. An overview of the measures obtained by our axioms is presented next.

2.1 The Preference for Improving Sequences of Outcomes

In our framework social welfare is a function of the distribution of individual trajectories of "well-being attributes" over time. A body of recent research has shown that, in a wide variety of choice situations, individuals prefer improving sequences of outcomes to declining ones that have comparable aggregate features (Loewenstein and Sicherman [1991], Frank and Hutchens [1993], Frederick, Loewenstein and O´Donoghue [2002]). This finding has proved to be particularly robust for sequences of monetary outcomes such as wage, income, and consumption profiles. Indeed, some of the studies show that people are willing to trade-off present income value in exchange for rising outcomes (Hsee, Abelson, and Salovey [1991]). In addition to monetary outcomes, preference for improving sequences has also been documented for certain health outcomes (Ross and Simonson [1991], Varey and Kahneman [1993], Chapman [2000]). A number of explanations have been offered to explain this pattern of choice which provides yet another piece of evidence against the commonly used discounted utility model. They include the anticipation of future well-being, commitment mechanisms, and debt aversion, among others.

Most of the studies above involve hypothetical "ex-ante" choices. A potential problem with this evidence is that hypothetical choices could be driven by a misperception of the actual well-being associated to the different type of trajectories.[10] In fact, mispredictions of future utility at the decision-making

[10]For example, some of this hypothetical choices could simply reflect the fact that individuals exhibit a tendency to like what they expect. Hence, if people expect rising

stage are well documented. For example, people tend to adapt more than they expect to changes in their circumstances.[11] Is it the case that improving sequences are associated with higher "experienced" (as opposed to "anticipated") well-being than declining sequences? Di Tella, Haisken-DeNew and MacCulloch [2007] find that changes in subjective well-being are consistent with loss aversion. Self-reported happiness adapts considerably less to negative income shocks than to comparable positive changes.[12]

In sum, the preference for rising rather than falling outcome profiles with commensurate aggregate outcomes is supported both by studies on ex-ante choices and evidence involving ex-post "experienced" well-being. We believe this individual-level evidence serves as a microfoundation for the *loss aversion axiom* presented in the sequel.

2.2 Poverty Status and Poverty Dynamics

A starting point of the poverty measurement literature is to identify those who are poor. In the simplest case, poverty is based on a single attribute of well-being, "income". In this unidimensional world, poverty is conceived as a condition or status associated with levels of income below an absolute threshold, the poverty line. The social welfare functional embodied by the poverty measure is then assumed to satisfy the focus axiom, which establishes that social welfare should only respond to the well-being of those who are poor.

The definition of a poverty status with multiple attributes of well-being is more delicate. Suppose for illustration that the poverty region is defined with respect to a vector of thresholds, one for each attribute. If an individual is above the threshold for some attributes but not others, should he or she be considered poor or not? Should we require individuals to below the poverty threshold in each attribute, one attribute alone or a subset of them? Assuming that a sensible choice is possible, the focus axiom can be applied to the corresponding poverty region just as in the case of unidimensional poverty measures. In practice, there is no obvious criterion to make this choice.

However, the problem of defining poverty is even more subtle in our case. Here multidimensionality arises from considering multiple periods even if

wages profiles over their productive lifes, these expectations could cause them to express a preference for rising wage profiles.

[11] See Kanheman (1997) and Gilbert (2005).

[12] Ongoing work by the authors confirms these findings.

a single per-period attribute such as consumption or income is considered. While the concept "poverty status" makes explicit reference to stable or static condition, in a dynamic setting welfare attributes may change over time. Instead of proposing a specific "dynamic" definition of poverty, we describe these welfare attributes in reference to an underlying static poverty definition. We adhere to the view that individuals may transition in and out of a static poverty condition. This has a number of advantages. First, from conceptual perspective, we avoid defining a poverty region which may be hard to justify and even at odds with the ontology of poverty status. Second, any static poverty measure induces a classification of individuals used to identify those who are dispossessed at a particular moment in time. In practice, this classification is often used to target policies and programs. Introducing a "dynamic" definition of poverty based on paths will necessarily lead to inconsistencies with respect to that classification.[13] Third, as will be clear later, the rankings produced by our measures will hold for any "dynamic" definition of poverty that takes deprivation at each period as a starting basis.

2.3 Overview of the Results

Consider an individual i that lives for two periods of time. At each period $t \in \{1, 2\}$ we observe a "well-being attribute" y_{it}, assumed to be a positive number. In the economics tradition, it is natural to think of y_{it} as consumption, but it could also be a measure of health, psychological well-being, or any dimension of human well-being.[14] Our method characterizes a family of multi-period measures of deprivation based on an underlying *static* definition of deprivation. We focus on static poverty measures with a (static) poverty line $z > 0$. The static measure takes the value of the attribute y_{it} as an input to determine individual deprivation $\delta_z(y_{it})$ at time t. The simplest static poverty measure is based on the poverty indicator function given by $\delta_z(y) = 1$ if $y \leq z$ and 0, otherwise. For this function, adding deprivation across all individuals in society, yields the head count measure of poverty. Another widely used poverty measure is the poverty gap, defined by the in-

[13] For example, if we decide that a "poor over time" is someone who is poor in one period and the government targets transfers to this group, it may provide transfers to someone extremely wealthy who escaped poverty.

[14] It is straighforward to extend the framework to a multidimensional vector of well-being attributes. This is done in section 4, which also discusses how to extend the framework to more than two periods.

dividual deprivation function $\delta_z(y) = \max\{0, 1 - \frac{y}{z}\}$, a normalized measure of the distance to the poverty line for those who are poor. Given any static deprivation function δ_z, we can associate to the stream $\mathbf{y}_i = (y_{i1}, y_{i2})$ a individual deprivation stream $\mathbf{d}_i$=(d_{i1}, d_{i2}) where $d_{it} = \delta_z(y_{it})$. For example, if δ_z is the poverty indicator, the deprivation stream $(0, 0)$ corresponds to an individual that was not poor in either period, $(0, 1)$ is the stream of someone who was not poor in period 1 but fell into poverty in period 2, and so on.

Our method associates a dynamic deprivation measure that takes the individual's deprivation stream $\mathbf{d}_i$ as an argument. An example used in the applications is the multi-period deprivation function

$$M(\mathbf{d}_i) = \frac{1}{2}(d_{i1} + d_{i2}) + \phi(d_{i2} - d_{i1}) \tag{1}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is given by

$$\phi(x) = \frac{1}{2}\left(\lambda \max\{0, x\} + \gamma \min\{0, x\}\right). \tag{2}$$

The function ϕ values "losses" or increases in deprivation ($x > 0$) at rate λ and "gains" or decreases in deprivation ($x < 0$) at a possibly different rate γ. Note that $\phi(0) = 0$, so paths associated with no change in deprivation over time have a dynamic deprivation equal to static deprivation in both periods. More generally, the measure implies that an individual's deprivation over time is an increasing function of absolute levels of deprivation at each period and also of changes in deprivation.[15] Aggregating across individuals we obtain the multi-period poverty measure

$$Q = \frac{1}{2}(\pi_1 + \pi_2) + \frac{1}{2}(\lambda L - \gamma G) \tag{3}$$

where π_t is the static poverty measure in period $t \in \{1, 2\}$,

$$L = \sum_{i:d_{i2}-d_{i1}>0} |d_{i2} - d_{i1}| \text{ and } G = \sum_{i:d_{i2}-d_{i1}<0} |d_{i2} - d_{i1}|.$$

[15] There is a parallel between this function and the preferences considered by Koszegi and Rabin [2006]. In their model, an agent's utility depends on absolute consumption levels c_1 and c_2 and also on changes in consumption relative to a reference point r. They consider a utility $u(c_1, c_2, r) = c_1 + c_2 + \phi(c_2 - r)$, where ϕ is a function satisfying prospect theoretic properties. The parallel with (1) is immediate if the reference is consumption in the first period, i.e., $r = c_1$.

Observe that Q combines information both about the poverty "stocks" as captured by the average static poverty term $\frac{1}{2}(\pi_1 + \pi_2)$, and poverty "flows" as captured by the term L associated to poverty creation and G associated to poverty destruction. For example, if the underlying static poverty measure is the poverty indicator then L is simply the number of people who fall into poverty in period 2 -those who lose- and G is the number of people who escape poverty in that period. As discussed later, for the poverty gap, the "flow" terms are, roughly, a weighted measure of individual income growth rates for the poor.

It is important to highlight two properties of (1) common to all the measures we consider. First, for fixed $d_{i2} - d_{i1}$, increases in deprivation at any period lead to higher dynamic deprivation. However, this does not imply that the multi-period measure is monotonic in d_1 and d_2. The measures we consider satisfy this additional monotonicity restriction. For the above example, this translates into constraints on the values of λ and γ. In particular monotonicity requires $|\gamma| \leq 1$ and $|\lambda| \leq 1$. Thus, monotonicity places an upper bound on the importance of "poverty flows" relative to "poverty stocks".

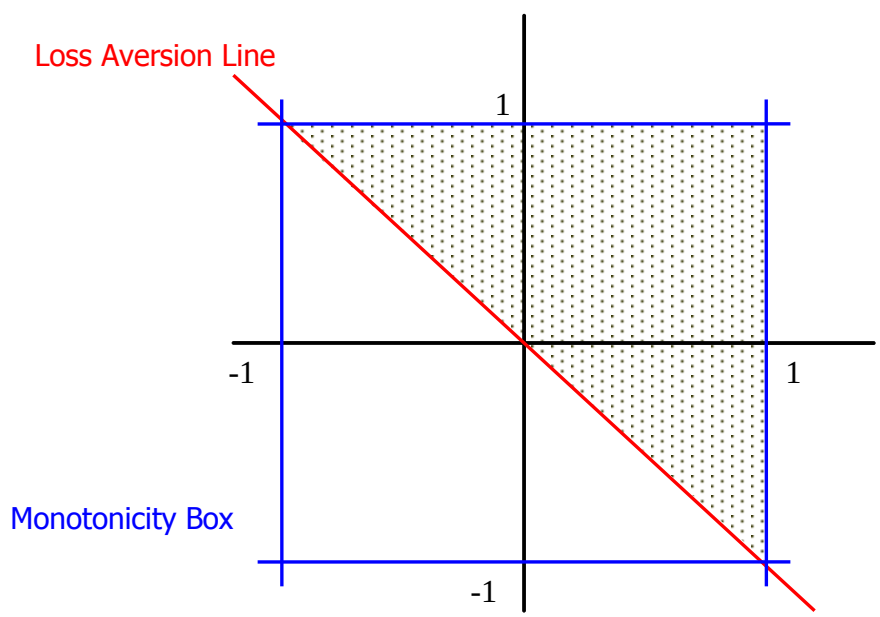

Figure 1: Monotonicity and Loss Aversion restrict weights on poverty

creation and destruction.

Our second key axiom is the loss aversion axiom, which is consistent with the evidence just discussed. To illustrate it, let $\mathbf{d}_i$=(d_{i1}, d_{i2}) and consider the deprivation path $\widehat{\mathbf{d}}_i$=(d_{i2}, d_{i1}), i.e., $\widehat{\mathbf{d}}$ "reverses" the deprivation stream $\mathbf{d}_i$. This means that, while the aggregate static deprivation across time is the same for both paths, if $\mathbf{d}_i$ is associated with an increase in deprivation, $\widehat{\mathbf{d}}_i$ is associated with decrease of the same magnitude. Note that

$$M(\mathbf{d}_i) - M(\widehat{\mathbf{d}}_i) = \frac{1}{2}(\lambda + \gamma)(d_{i2} - d_{i1})$$

so that, if $\lambda + \gamma > 0$, the sign is positive for an increase in deprivation and negative for decrease.[16]

Figure 1 illustrates the constraints imposed by our axioms -monotonicity and loss aversion- in the two-dimensional space (λ, γ) of parameters that define the family of linear measures of poverty dynamics introduced above. Measures satisfying both of these axioms are associated with the shaded triangle.

We conclude this section with a numerical example that illustrates the order induced by the set of measures just described on the space of income dynamics. Table 1a below summarizes the distribution of income streams for individuals in three hypothetical societies A, B and C. The poverty line is assumed to be 10 monetary units, and incomes above this threshold are coded as 10^+. For simplicity, we assume that there are four categories of income trajectories, one corresponding to those who are never poor, one corresponding to those who are always poor (but who still experiment an income rise from 5 to 7.5), individuals who fall into poverty, and individuals who enter poverty. The first three columns of the table describe the distribution of income streams for each society. For example, in society A 88% of the population is never poor and so on. The last two columns of Table 1a presents the deprivation streams for each of the corresponding income trajectories. Each of these columns uses one of the underlying static poverty measures introduced above, the poverty indicator and the poverty gap, respectively.

[16] Observe that this restriction does not necessarily imply that gains will be associated with a decrease in deprivation, which requires $\gamma > 0$. While this is certainly a possibility, the framework allows for upwards "adjustment costs" but it requires these costs to be smaller for poverty outflows than poverty inflows.

Table 1a: *Numerical Example: Income and Deprivation Streams*

Society A	Society B	Society C	(y_1, y_2)	(d_1, d_2) Indicator	(d_1, d_2) Gap
0.88	0.80	0.85	$(10^+, 10^+)$	$(0, 0)$	$(0, 0)$
0.06	0.10	0.02	$(5, 10^+)$	$(0, 1)$	$(0, 0.5)$
0.03	0.05	0.05	$(10^+, 5)$	$(1, 0)$	$(0.25, 0)$
0.03	0.05	0.08	$(5, 7.5)$	$(1, 1)$	1

Table 1b: *Numerical Example: Poverty Stocks and Flows*

	Society A	Society B	Society C
1. *Poverty Indicator*			
Static poverty at $t = 1$: π_1	0.09	0.15	0.10
Static poverty at $t = 2$: π_2	0.06	0.10	0.13
Poverty Creation: L	0.03	0.05	0.05
Poverty Destruction: G	0.06	0.10	0.02
2. *Poverty Gap*			
Static poverty at $t = 1$: π_1	0.0450	0.0750	0.0500
Static poverty at $t = 2$: π_2	0.0225	0.0375	0.0450
Poverty Creation: L	0.0150	0.0250	0.0250
Poverty Destruction: G	0.0375	0.0625	0.0300

Using Table 1a it is straightforward to compute static poverty π_t at each period t, poverty creation L and poverty destruction G. This is reported in table 1b. These are the inputs required to compute the multi-period poverty measure of formula (3). For example, taking the poverty gap as the underlying static poverty measure, the multi-period poverty measures for each society as a function of λ and γ are given by

$$\begin{aligned} Q_A(\lambda, \gamma) &= 0.03375 + \lambda 0.015 - \gamma 0.0375 \\ Q_B(\lambda, \gamma) &= 0.05625 + \lambda 0.025 - \gamma 0.0625, \text{ and} \\ Q_C(\lambda, \gamma) &= 0.0475 + \lambda 0.025 - \gamma 0.0300. \end{aligned}$$

For a fixed pair (λ, γ) these indexes define a ranking between societies. The family of *all* measures consistent with monotonicity and loss aversion induces a partial order: the income dynamics of society X dominates the income

dynamics of society Y if $Q_X(\lambda,\gamma) \leq Q_Y(\lambda,\gamma)$ for *all* pairs (λ,γ) in the shaded triangle of Figure 1. From above, we have that

$$\Delta Q_{AB}(\lambda,\gamma) \equiv Q_A - Q_B = -0.01375 - \lambda 0.01 + \gamma 0.025,$$

which can be shown to be strictly negative for all λ and γ.[17] Hence, society A dominates society B. Note that poverty reduction $\pi_2 - \pi_1$ is higher in society B and so is "poverty mobility" as measured by $L + G$. However, our measures weigh stocks, and average poverty in society A is considerably lower than in society B. Further, even though society B has considerably more poverty destruction than A it also has more poverty creation, and losses are more important than gains for indexes consistent with loss aversion. A similar exercise shows that A also dominates C.

Comparing B and C we see that

$$\Delta Q_{BC}(\lambda,\gamma) \equiv Q_B - Q_C = 0.00875 - \gamma 0.0325.$$

This number is positive if $\gamma \leq 13/35 \simeq 0.27$ and positive otherwise. Thus, for some measures in the family B is better than C and for others the ranking is reversed. In particular, if poverty flows are valued relatively more, society B is ranked better than society C and vice versa. Indeed, society B has considerably higher average poverty than C but also more poverty destruction.

This simple example illustrates that, in general, the ranking induced by our measures will be different than the ones induced by looking at poverty reduction and poverty mobility. We further discuss this in light of our empirical applications. In the sequel we present the main axioms that define our multi-period poverty measures consistent with a preference for increasing sequences of outcomes given the same aggregate deprivation across time. We then characterize the order induced by these measures over the space of income dynamics.

3 Multi-period Poverty Measures

We introduce poverty measures based on a multi-period stream of welfare attributes. Society consists of a cohort of N individuals. Each time period

[17] Since $\Delta Q_{AB}(\lambda,\gamma)$ is linear in the parameters, minimizing over those pairs (λ,γ) that are in the shaded triangle is a linear program. The minimum is thus achieved by one of the vertexes of the triangle and evaluating the objective at these vertices shows that $min \Delta Q_{AB}(\lambda,\gamma) < 0$.

or "time window" is denoted $t \in \{1, ..., T\}$.[18] For short, the welfare attribute is referred as "income". The income of individual i at time period t is denoted by y_{it} and $\mathbf{y}_i = (y_{i1}, y_{i2}, ...y_{iT})$ is individual i's *income path*. The set of possible income levels is denoted by Y. We assume that Y is endowed with a complete linear order denoted by $\leq$. In particular this allows for a categorical welfare attribute such as the employment status of an individual.[19] The set of all possible income paths is Y^T. At time t the cohort has a profile of incomes $\{y_{it}\}_{i=1}^N \in Y^N$. The latter is equivalent to a static income distribution. A profile of income paths for the cohort is $\{\mathbf{y}_i\}_{i=1}^N \in Y^{TN}$. We focus on multi-period poverty measures based on transitions in and out of a poverty status. This status is based on an underlying static (per period) definition and measure of poverty.

3.1 Poverty Status and Underlying Static Poverty

The static *poverty line* is an income level $z \in Y$ and an individual $i \in N$ is said to be *poor at time* t if $y_{it} \leq z$. Static poverty measures are based on an *individual measure of deprivation* and *aggregation* across individuals. The individual measure of deprivation satisfies a set of "core" axioms: focus, monotonicity, continuity, and normalization. The main axioms for aggregation are symmetry and subgroup decomposability.[20] The static measures considered hereafter satisfy these axioms as well. The properties implied by these axioms are summarized by the following definition.

Definition 1 (Static Measure) *A function $\delta_z : Y \to \mathbb{R}$ is an individual static measure of deprivation or poverty if it is non-increasing, $\delta_z(0) = 1$ and*

[18] In our empirical applications each "time window" corresponds to five calendar years. Hence, time periods should not be confounded with calendar years. The number of calendar observations per window is dictated by the underlying definition of poverty. In our applications, poverty status is defined on the basis of five years of income observations.

[19] As illustrated by our applications our method requires longitudinal data. In many developing countries panel data on relevant welbeing attributes are hard to find but it might be possible to construct past values of a stream by required the subjects surveyed for information from previous years. In this case, categorical data such as the employment status might be considerably less noisy than other measures.

[20] See the Appendix for a formal defintion of the axioms. Subgroup decomposability implies that the poverty measure is additive across members of society. Symmetry implies that the poverty index levels of the variables that define poverty depend exclusively on the level of income of a particular individual rather than her idenitity, thus the same individual deprivation function is used across individuals.

$\delta_z(y) = 0$ *for all* $y > z$. *A function* $\pi : Y^N \to \mathbb{R}$ *is said to be an admissible static poverty measure if*

$$\pi(\{\mathbf{y}_{it}\}) = \frac{1}{N} \sum_{i \in N} \delta_z(y_{it}).$$

for some individual deprivation function δ_z. *The set of all admissible static poverty measures is denoted by* Π.

We write $\pi[\delta_z]$ to designate the static poverty measure with individual deprivation function δ_z. For example, if $1(\cdot)$ is the indicator function and

$$\delta_z^0(y) = 1(y \leq z)$$

then $\pi[\delta_z^0]$ is the headcount ratio, the share of individuals that are poor at a given moment in time. A popular family of static poverty measures introduced by Foster, Greer and Thorbecke [1984] (FGT) is defined by a single parameter $\alpha \geq 0$ and the individual deprivation function

$$\delta_z^\alpha(y) = \delta_z^0(y) \left| 1 - \frac{y}{z} \right|^\alpha.$$

For $\alpha = 0$, the reduces to the poverty indicator. For $\alpha > 0$, the formula is sensitive to the distance to the poverty line. The case of $\alpha = 1$ corresponds to the widely used poverty gap.

For simplicity we consider an additional restriction on static poverty measures. In particular, we assume that unless the poverty measure is the headcount ratio ($\delta_z = \delta_z^0$), individual deprivation is strictly monotonic in the poverty region. Fix a static deprivation measure δ_z and let $D[\delta_z] = \delta_z(Y)$ be the set of possible deprivation values and write $D = D[\delta_z]$ whenever it leads to no confusion. Our previous assumption, implies that either $D = \{0, 1\} = \delta_z^0(Y)$ or else $D = [0, 1]$ as is the case for all the FGT family when $\alpha > 0$.[21]

3.2 Multi-period Deprivation and Poverty Dynamics

We define measures of poverty associated to income paths that satisfy a similar set of basic axioms than the ones assumed for static measures.[22]

[21] This is a mild restriction and is mostly for exposition.

[22] In this context, intertemporal deprivation is determined by two attributes, incomes at period 1 and 2. The same axioms used to derive static income based poverty measures and can be used to derive multidimensional poverty measures. See for example Bourguignon and Chakravarty [2003].

Definition 2 *A function $q : Y^T \to \mathbb{R}$ is an individual measure of multi-period poverty if it is non-increasing, its minimum value is zero and its maximum value is 1. A function $Q : Y^{TN} \to \mathbb{R}$ is said to be an admissible multi-period poverty measure if*

$$Q(\{\mathbf{y}_i\}) = \frac{1}{N} \sum_{i \in N} q(\mathbf{y}_i). \tag{4}$$

Just as in the case of static poverty measures, the above definition assumes that the individual multi-period poverty measures satisfy *monotonicity* and *normalization*. Similarly, the aggregate poverty measure can be derived from the subgroup decomposability and the symmetry axioms. However, in contrast to static poverty measures we do *not* require the existence of a multi-period poverty region. The above definition is agnostic about identifying perfectly those who are "poor over time". As argued below, we are ultimately interested in the orders on income processes induced by a family of poverty measures and, these poverty measures should allow for any reasonable multidimensional definition of poverty over time. In what follows we provide axioms that restrict the nature of the individual multi-period deprivation measure q.

4 Main Axioms and Representation

We introduce an axiomatic foundation for a family of measures that generalize the one introduced in section 3. Our main result is an explicit representation characterizing the individual multi-period measures that satisfy it. All proofs are in Appendix.

In the example of section 3, the individual multi-period deprivation measures are allowed to depend both on levels and changes of an index of deprivation or disutility over time. This index serves as the unit measure to evaluate the well-being at each period and its changes -gains and losses- over time. Our scaling axiom introduces this measure.

Axiom (SC) [Scaling] *For each q there exists an static deprivation index δ_z such that $q(\mathbf{y}) = M(\delta_z(y_1), ..., \delta_z(y_T))$ for some monotonic function $M : D^T \to [0, 1]$, where $D = \delta_z(Y)$.*

A few remarks are in place. First, we use $q[M, \delta_z]$ to designate a measure that satisfies axiom (SC) for a given scale δ_z and a monotonic function

M. Second, given δ_z a measure q is entirely determined by the monotonic transformation M. Further for a fixed δ_z, any income path $\mathbf{y} = (y_1, ..., y_T)$ is associated with a deprivation path $\mathbf{d} = (\delta_z(y_1), ..., \delta_z(y_T))$. All of our axioms can be restated in terms of deprivation paths for a fixed δ_z and interpreted as restrictions on the space of monotonic functions M. We have chosen to present the axioms in terms of incomes to emphasize the primitive. Third, from theoretical point of view, axiom does not prevent the measure to depend directly on levels or paths of the welfare attributes y_t. For example, $Y = \mathbb{R}$ δ_z is linear (as in the poverty gap), and the poverty line z can be chosen to be arbitrarily large. The axiom allows for much more general disutility indexes, it accommodates an arbitrary increasing function δ_z and z.. From a practical point of view, however, we find it useful to focus on commonly static poverty measures. In this context, axiom (SC) says that the poverty measure keeps track of transitions in and out of a pre-defined static poverty status.

To introduce our next axioms, we point out that a deprivation path $\mathbf{d} = (d_1, ..., d_T)$ can always be described by the $(T-1)$-component vector of deprivation changes $w(d) = (d_2 - d_1, ..., d_T - d_{T-1})$, and a real-valued map $\nu(d)$ that is strictly increasing in all of its arguments. Formally, for any such $\nu : D^T \to \mathbb{R}$, the map that assigns to each $\mathbf{d}$ the pair $(w(\mathbf{d}), \nu(\mathbf{d}))$ is a bijection. Furthermore, the map ν can be chosen to depend exclusively on income levels regardless of their sequence in the stream. To make this precise, we introduce some notation. Each income stream $\mathbf{y} \in Y^T$ induces a probability distribution $\sigma(\mathbf{y}) \in \Delta(Y)$ on income levels: for each $\widehat{y} \in Y$ the distribution $\sigma(\mathbf{y})$ assigns the "empirical" frequency $\sigma(\mathbf{y})(\widehat{y}) \equiv \sum_{t=1}^{T} \chi(y_t = \widehat{y})/T$ to income $\widehat{y}$. The distribution $\sigma(\mathbf{y})$ on Y induced by $\mathbf{y}$ is invariant to permutations of the sequence of incomes over time periods.[23] The distribution induced by a path is entirely independent of the trajectory or income changes between consecutive periods. Thus, if $q(\mathbf{y})$ is a function of $\sigma(\mathbf{y})$ alone it is independent of income or deprivation changes. On the other hand, given an index δ_z and an income path $\mathbf{y}$, the $(T-1)$-dimensional vector of deprivation changes is

$$\omega(\mathbf{y}; \delta_z) = (\delta_z(y_2) - \delta_z(y_1), ..., \delta_z(y_T) - \delta_z(y_{T-1})).$$

Our next axiom imposes a key separability condition.

[23] For example, if $T = 2$, $\mathbf{y} = (10, 20)$ and $\mathbf{y}' = (20, 10)$, then $\sigma(\mathbf{y}) = \sigma(\mathbf{y}')$, and this distribution assigns probability $1/2$ to both 10 and 20 regardless of their sequence in time (and 0 to any other $y \in Y$).

Axiom (SF) [Stocks-Flows Separability] *There exists a function* $\mu : Y^T \to \mathbb{R}$ *increasing in all of its arguments (and strictly so for* $y_t < z$) and *such that*

(SF1) $\sigma(\mathbf{y}) = \sigma(\mathbf{y}') \Rightarrow \mu(\mathbf{y}) = \mu(\mathbf{y}')$

(SF2) For all $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}' \in Y^T$ *such that* $\mu(\mathbf{x}) = \mu(\mathbf{y})$, $\mu(\mathbf{x}') = \mu(\mathbf{y}')$, $\omega(\mathbf{y};\delta_z) = \omega(\mathbf{y}';\delta_z)$ *and* $\omega(\mathbf{x};\delta_z) = \omega(\mathbf{x}';\delta_z)$

$$q(\mathbf{x}) \geq q(\mathbf{x}') \Rightarrow q(\mathbf{y}) \geq q(\mathbf{y}').$$

We note that condition $(SF1)$ vacuous on its own since, as just argued, we can always decompose a stream into a component that keeps track of changes from one period to the next and a "residual" that is invariant to permutations of income levels across periods. However, combined with $(SF2)$, the residual must be such that a deprivation measure can be expressed as a separable function of $\mu(\mathbf{y})$ and $\omega(\mathbf{y}';\delta_z)$. Indeed, in the Appendix we show that axiom (SF) implies the existence of a monotonic function $S : \mathbb{R} \to \mathbb{R}$ and a function $\Phi : U \to \mathbb{R}$ such that $q(\mathbf{y}) = S(\mu(\mathbf{y})) + \Phi(\omega(\mathbf{y}';\delta_z))$, where $U = \{w \in [-1,1] | w = d - d',\ d, d' \in D\}$ is the set of possible deprivation changes from one period to the next.. Intuitively, we can think of $S(\mu(\mathbf{y}))$ as valuation of income levels that is independent of the "shape" of the stream. Thus, (SF) is a strong restriction as it makes the valuation of deprivation changes -"flows" independent of the valuation of deprivation levels -"stocks" as captured by $\mu(\mathbf{y})$. At the same time, this stock-flow separability yields an intuitive and is shared by prominent examples in the recent literature on reference-dependent preferences over consumption streams (Gilboa [1989], Bowen, Minehart, and Rabin [1999], Koszegi and Rabin [2006]). We view it as intuitive benchmark that, as shown briefly, allows for an explicit characterization of dominance relationship on the space of income stream distributions. Note also that, combined with axiom (SC), we must have that $\mu(\mathbf{y}) = \nu(\delta_z(y_1), ..., \delta_z(y_T))$ for some function ν that is also invariant to permutations of its arguments. In our motivating example, $T = 2$, $\nu(d_1, d_2) = (d_1 + d_2)/2$ and $\Phi(w) = \lambda \max\{0, w\} + \gamma \min\{0, w\}$.

We need some notation to introduce our next axiom. Let $R : Y^T \to Y^T$ denote the "reflection map" defined by $R(y_1, y_2, ..., y_T) = (y_T, ..., y_2, y_1)$. We write

$$I(\underline{y}, \overline{y}) = \{\mathbf{y} \in Y^T | y_1 = \underline{y}, y_T = \overline{y}, \text{and } y_{t+1} \geq y_t, y_t \in \{\underline{y}, \overline{y}\}\}$$

for the set of increasing incomes streams with "support" $\{\underline{y}, \overline{y}\}$, i.e., at each period income is either $\underline{y}$ or $\overline{y}$. Observe that if $\mathbf{y} \in I(\underline{y}, \overline{y})$ is a path of

increasing incomes then $R(\mathbf{y})$ is a path of declining incomes.[24] Note also that $\sigma(\mathbf{y}) = \sigma(R(\mathbf{y}))$, i.e., $\mathbf{y}$ and $R(\mathbf{y})$ are associated with the same distribution on $\{\underline{y}, \overline{y}\}$. Hence, while each of these paths is associated with exactly the same number of periods of incomes $\underline{y}$ and $\overline{y}$, $\mathbf{y}$ is associated with a gain over time, whereas $R(\mathbf{y})$ is associated with a loss.

Axiom (LA) [Loss Aversion] *For any $\underline{y}, \overline{y} \in Y$ with $\underline{y} \leq \overline{y}$ and any income stream $y \in I(\underline{y}, \overline{y})$, the individual deprivation measure of multi-period poverty $q : Y^T \to [0, 1]$ satisfies $q(\mathbf{y}) \leq q(R(\mathbf{y}))$, and $q(R(\mathbf{y})) - q(\mathbf{y})$ weakly increases as $\overline{y} - \underline{y}$ increases.*

In words, in line with the evidence discussed in section 2, axiom (LA) expresses the idea that an individual experiences (weakly) higher well-being from an increasing path than a decreasing one.

Our next two axioms restrict the dependency of our measures on trajectories when $T \geq 3$. Let $r \geq 1$ and $s \leq T$ and denote by $A = [r, s]$ the interval of times $\{r, r+1, ..., s\}$. Given two arbitrary income streams $\mathbf{x}, \mathbf{k} \in Y^T$ we use $(\mathbf{k}_A, \mathbf{x}_{\sim A})$ to denote the income stream $\mathbf{x}' \in Y^T$ such that $x'_t = k_t$ for $t \in A$ and $x'_t = x_t$ for $t \in T \backslash A$.

Axiom (TD) [Time Decomposability] *For any non-empty time interval $A = [r, s]$ and income streams $\mathbf{x}, \mathbf{y}, \mathbf{k} \in Y^T$ such that*

$$x_{r-1} = y_{r-1} \text{ and } x_{s+1} = y_{s+1}$$

the multi-period deprivation measure $q : Y^T \to [0, 1]$ satisfies

$$q(\mathbf{k}_A, \mathbf{x}_{\sim A}) \geq q(\mathbf{k}_A, \mathbf{y}_{\sim A}) \Leftrightarrow q(\mathbf{k}'_A, \mathbf{x}_{\sim A}) \geq q(\mathbf{k}'_A, \mathbf{y}_{\sim A})$$

for any $\mathbf{k}' \in Y^T$.

The axiom says that the ranking between two income streams that coincide on an interval of times A and that are also identical in the time periods that border it must be determined by their values outside this interval. It

[24] We remind the reader that, in keeping with the poverty literature, in this paper, "income" refers simply the relevant welfare attribute. Perhaps consumption would correspond more precisely to what most economist would consider a natural carrier of utility and, in this line, a declining (increasing) could be thought of as a path of declining (increasing) consumption.

captures a limited notion of time-separability of individual multi-period deprivation. Indeed, the axiom places no restriction on intervals that coincide on A but not on the borders of A. Without this restriction a standard additive representation of q would obtain. Instead, Lemma 5 in the Appendix shows that (TD) implies that q can be represented as $q(\mathbf{y}) = \sum_{t=1}^{T-1} q_{t+1}(y_t, y_{t+1})$, i.e., a sum of period-deprivation functions that can depend on previous period income.[25] The axiom is appealing for at least two reasons. First, it allows for dependence on trajectories, i.e., for complementarity between income values across different periods. Second, it limits these complementarities to consecutive periods. In sum, it allows deprivation to depend on levels and changes from one period to the next but not on more global properties of the income stream. In general, it might reasonable to expect well-being on other characteristics of the income path. This may ultimately depend on the choice of a time window. Further, the fact that the measure is only a function of consecutive periods allows the analyst to make a welfare assessment as soon as a new time period observation is added, something important in practice.

Axioms (SF) and (TD) jointly imply the following:

Lemma 1 *The admissible (individual) multi-period poverty measure $q[M, \delta_z]$ satisfies axioms (TD) and (SF) if and only if there exist $m : D \to [0,1]$ with $m(0) = 0$ and $m(1) = 1$, and flow-value functions $\phi_t : U \to \mathbb{R}$ with $\phi_t(0) = 0$, $t \in \{1, ..., T-1\}$ such that*

$$M(\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} m(d_t) + \frac{1}{T}\sum_{t=1}^{T-1} \phi_t(d_{t+1} - d_t) \tag{5}$$

for each $\mathbf{d} = (d_1, d_2, ..., d_T)$.

Thus, M has two components, one that evaluates deprivation levels and other one that evaluates changes in deprivation. Our final axiom requires the valuation of changes in well-being over time to remain constant over life-cycle. That is, the multi-period deprivation value of a gain (loss) is independent of whether this change takes place early or late along the income path. This means that the measure of multi-period deprivation is symmetric

[25] This limited-time separability is also shared by the by the preferences in Gilboa [1989] and Koszegi and Rabin [2006]. The (TD) axiom is weakening of Gilboa's Variation-Preserving Sure Thing Principle axiom.

with respect to time-periods, something referred as *calendar neutrality*. To avoid duplicating notation, we state the axiom as a strengthening of the (SF) axiom. Each vector of deprivation changes $\mathbf{w} = (w_1, ..., w_{T-1})$ induces a distribution $\sigma_w(\mathbf{w})$ on the set of possible deprivation changes U such that where the probability of each $\widehat{w} \in U$ is simply the "empirical frequency" $\sigma_w(\mathbf{w})(\widehat{w}) = \sum_{t=1}^{T-1} \chi(w_t = \widehat{w})/(T-1)$ that $\mathbf{w}$ assigns to it.

Axiom (CN) [Calendar Neutrality] *For all* $\mathbf{x}, \mathbf{y} \in Y^T$ *such that* $\sigma(\mathbf{x}) = \sigma(\mathbf{y})$ *and* $\sigma_w(\omega(\mathbf{x};\delta_z)) = \sigma_w(\omega(\mathbf{y};\delta_z))$ *we have that* $q(\mathbf{x}) = q(\mathbf{y})$.

In words, the (CN) axiom says that income streams associated with the same distribution of income levels σ and also the same distribution of deprivation changes σ_w yield the same multi-period deprivation. Combined with Lemma 1 we obtain the following:

Lemma 2 *Suppose the admissible (individual) multi-period poverty measure* $q[M, \delta_z]$ *satisfies axioms* (TD) *and* (SF). *If, in addition,* $q[M, \delta_z]$ *satisfies axiom* (CN) *then there exists* $\phi : U \to \mathbb{R}$ *such that (5) is satisfied with* $\phi_t = \phi$ *for all* $t \in \{2, ..., T\}$.

Thus, the framework allows to incorporate deprivation measures that account for life-cycle adjustments in the valuation of deprivation changes by relaxing the (CN) axiom. It might be sensible to consider different deprivation standards as the individual ages.

Lemmas 1 and 2 provides us with a representation of M that is entirely determined by a pair of functions (m, ϕ), where the first component values deprivation levels and the second one values changes. We write $M = M(m, \phi)$. However, this representation does not incorporate two of our key properties, the monotonicity of M and the (LA) axiom.

Definition 3 *Let* $m : D \to [0, 1]$ *and* $\phi : U \to \mathbb{R}$. *The pair* (m, ϕ) *is said to respect monotonicity if for any* $d, d' \in D$ *such that* $d' \geq d$ *and any* $d_+, d_- \in D$ *we have that*

$$\begin{aligned} (M1)\ \phi(d_+ - d) - \phi(d_+ - d') &\leq m(d') - m(d) \\ (M2)\ \phi(d' - d_-) - \phi(d - d_-) &\leq m(d') - m(d) \end{aligned}$$

and, if $T \geq 3$ *then*

$$(M3)\ \phi(d_+ - d) - \phi(d_+ - d') - (\phi(d' - d_-) - \phi(d - d_-)) \leq m(d') - m(d).$$

The conditions are equivalent to the monotonicity requirement on $M(m, \phi)$. Observe that if m and ϕ are differentiable the (M1) and (M2) translate into $|\phi'(w)| \leq 1$ for each $w \in [-1, 1]$, and if $T \geq 3$, (M3) is $|\phi'(w) - \phi'(v - w)| \leq 1$ for each $v, w \in [-1, 1]$. Intuitively, monotonicity restricts the contribution of flows to deprivation relative to the contribution of stocks. In our motivating example it constrains the (absolute) values of γ and λ -which measure the importance of deprivation changes relative to levels, to be bounded above by one.

Similarly, the (LA) axiom constrains the flow value function ϕ. In particular, it distinguishes between the contributions of losses and gains to multi-period deprivation.

Definition 4 *The flow function ϕ satisfies the loss aversion (LA) property if for each $w \geq 0$ we have that*

$$(LA)\ \phi(w) \geq \phi(-w) \text{ and } \phi(w) - \phi(-w) \text{ increases with } w.$$

The property says that, in absolute value, the contribution of losses in well-being of multi-period deprivation is larger than the contribution of gains, and that this difference increase with the size of the change. Note that the (LA) property allows both for a gain in well-being to contribute or reduce intertemporal deprivation. The first case can be interpreted an adjustment cost regardless of the direction of change (as in Gilboa [1989]). A stronger version of the (LA) axiom is required to produce the more restrictive version of loss aversion in which gains are associated with a positive flow of utility, losses with a negative flow, and "losses loom larger than gains".[26]

A complete characterization of the family of measures that satisfy all of the axioms above is summarized by the Theorem below.

Theorem 1 *The admissible (individual) multi-period poverty measure $q[M, \delta_z]$ satisfies axioms (SF), (LA), (TD), and (CN) if and only if there exist a pair*

[26] In particular, this is exactly what obtains if we strengthen the (LA) axiom as follows: for the set paths with support $\{\underline{y}, \overline{y}\}$ that have have the same distribution of incomes (i.e., the same number of periods at $\underline{y}$ and $\overline{y}$) the path associated with an increasing income is stream (i.e., $y_t = \underline{y}$ for $t \leq n$ and $y_t = \overline{y}$ for $t > n$) is the one associated withe the lowest deprivation. This strengthening, implies that for $v > 0$, ϕ satisfies $\phi(v) > 0$ for and $\phi(v) \geq -\phi(-v)$.

of functions $m : D \to [0,1]$ *with* $m(0) = 0$ *and* $m(1) = 1$*, and* $\phi : U \to \mathbb{R}$ *with* $\phi(0) = 0$ *such that*

$$M(\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} m(d_t) + \frac{1}{T}\sum_{t=1}^{T-1} \phi(d_{t+1} - d_t) \tag{6}$$

for each $\mathbf{d} = (d_1, d_2, ..., d_T)$*. Furthermore, the flow-value function* ϕ *satisfies the (LA)-property and* (ϕ, m) *satisfies the monotonicity restrictions (M1)-(M3).*

Theorem 1 says that any deprivation function $q[M, \delta_z]$, consistent with the (LA) can be expressed as the sum of term $S(\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} m(d_t)$ that is insensitive to the trajectory and depends exclusively on the levels of deprivation experienced at each period regardless of the order in the sequence, and a function $\Phi(\mathbf{d}) = \sum_{t=1}^{T-1} \phi(d_{t+1} - d_t)$ that measures the value of flows and captures the preference for increasing sequences. In addition, an admissible deprivation function $M(\cdot)$ must be monotonic. Comparing (1) and (6), the function M is generalizes the linear multi-period measure presented in section 2.3 as m and ϕ are not piece-wise linear.

We conclude this section mentioning the fact that while our characterization has focused on dynamics that track the evolution of a single welfare attribute, it is relatively straightforward to extend the framework multiple well-being attributes if the corresponding static multi-attribute poverty region and measure has been defined.

5 Ranking Poverty Dynamics

In this section we provide a complete characterization of the (partial) ranking induced by our measures on the space of income stream profiles or distributions. We start with some basic notation and definitions. Fix underlying static deprivation measure δ_z and recall that $D \equiv \delta_z(Y)$. Given δ_z we can identify a profile of income streams $\{\mathbf{y}_i\} \in Y^{TN}$ with the profile of deprivation streams $\{\mathbf{d}_i\} \in D^{TN}$ where each $\mathbf{d}_i = (\delta_z(y_{i1}), ..., \delta_z(y_{iT})) \in D^T$.

Definition 5 (Poverty Dynamics) *Fix* δ_z *and let* $\{\mathbf{d}_i\} \in D^{TN}$ *be a profile of deprivation streams. The poverty dynamics associated to* $\{\mathbf{d}_i\}$ *is the distribution* $f \in \Delta(D^T)$ *such that* $f(\mathbf{x}) = \frac{1}{N}\sum_{i \in N} 1(\mathbf{d}_i = \mathbf{x})$ *for each* $\mathbf{x} \in D^T$*.*

The distribution f summarizes the realized deprivation paths for all individuals in the cohort of N individuals. Recall that, by assumption, we either have that $D = \{0, 1\}$ -binary deprivation levels associated to the poverty indicator ($\delta_z = \delta_z^0$), or else $D = [0, 1]$ (δ_z is strictly increasing and continuous on Y). In the latter case, the space of all possible poverty dynamics $\Delta(D^T)$ is independent of δ_z. Of course, the same profile of income streams will be associated with different elements of $\Delta(D^T)$ as we vary δ_z.

For this fixed δ_z an individual deprivation measure q consistent with our axioms is entirely determined by a map $M : D \to [0, 1]$ that satisfies the conditions of Theorem 1. The set of all such functions is denoted $\mathcal{M}$. With some abuse of notation, whenever it leads to no confusion, we write $q[M]$ instead of $q[M, \delta_z]$ for the individual poverty measure induced by $M \in \mathcal{M}$ given δ_z. Similarly, we use $Q[M]$ for the aggregate measure associated to $q[M]$ and, for this measure, multi-period poverty for a society described by f is given by

$$Q[M](f) = \int_{\mathbf{x} \in D^T} M(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

Each family of multi-period poverty measures induces a partial order on $\Delta(D^T)$, the space of poverty dynamics:

Definition 6 (Multiperiod Poverty Order) *Let $\mathcal{Q}$ a set of multi-period poverty measures. Given two poverty dynamics $f^A, f^B \in \Delta(D^T)$, we say that f^A dominates f^B for the set $\mathcal{Q}$ if $Q(f^A) \leq Q(f^B)$ for any $Q \in \mathcal{Q}$. Whenever this holds we write $f^A \succeq f^B$.*

In the sequel we characterize partial order on induced by the family $\mathcal{Q}^* = \{Q[M] | M \in \mathcal{M}\}$ of multi-period measures satisfying our axioms.

Remark 1 ***The Dominance Optimal Control Problem.***

Let $K[M](f^A, f^B) \equiv Q[M](f^A) - Q[M](f^B)$ and

$$K^{\max}(f^A, f^B) \equiv \sup_{M \in \mathcal{M}} K[M](f^A, f^B). \qquad \text{(DOCP)}$$

Note that $f^A \gtrsim f^B \Leftrightarrow K^{\max}(f^A, f^B) \leq 0$. Hence, characterizing the dominance relation $\gtrsim$ is equivalent to characterizing the value of the optimization problem (DOCP).

Before turning to a characterization the solution of (DOCP) we introduce some notation. Let $\mathbf{u}_{-t} = (u_1, ..., u_{t-1}, u_{t+1}, ..., u_T) \in D^{T-1}$ and write $f_t(x) = \int_{D^{T-1}} f(u_1, ..., u_{t-1}, x, u_{t+1}, ..., u_T) d\mathbf{u}_{-t}$ for the period-t *marginal distribution of deprivation levels* associated to poverty dynamics f. Observe that static poverty at time t is $\pi_t = \int_D x f_t(x) = 1 - \int_D F_t(x)$, where F_t is the cumulative distribution. Denote by

$$\overline{f}(x) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t(x)$$

the time-average of these marginal distributions. Thus, the time-average of static poverty over periods 1 through T is $\overline{\pi} = \int_D x \overline{f}(x) = 1 - \int_D \overline{F}(x)$.

Similarly let

$$f_{t,t+1}(x, x') = \int_{D^{T-1}} f(u_1, ..., u_{t-1}, x, x', u_{t+2}, ..., u_T) du_1 ... du_{t-1} du_{t+2} ... du_T$$

be the density of deprivation paths with deprivation levels $d_t = x$ and $d_{t+1} = x'$ in periods t and $t+1$ respectively, associated to the poverty dynamics f. We can use this density to compute the distribution of deprivation changes between these time periods. Recall that $U = \{v | \ v = x' - x, \ x, x' \in D\}$ is the set of all possible deprivation changes. Let $C(v) = \{(x, x') \in D^2 | x' - x = v\}$ be the set of deprivation pairs that lead to a deprivation change $v \in U$. Given $f_{t,t+1}$ the *distribution of deprivation changes from t to $t+1$* induced by f is given by

$$h_{t,t+1}[f](v) = \int_{C(v)} f_{t,t+1}(x, x') dx dx'$$

for each $v \in U$. We define the average (normalized) distribution of deprivation changes induced by f as integral

$$h_f(u) = \frac{1}{T} \sum_{t=1}^{T-1} h_{t,t+1}[f](u).$$

Lemma 3 *Suppose that M satisfies (6) for the pair (m, ϕ). Then*

$$Q[M](f) = \int_0^1 m(x) \overline{f}(x) dx + \int_{-1}^1 \phi(v) h_f(v) dv.$$

The result uses Theorem 1 and shows that we can express the aggregate multi-period deprivation $Q[M](f)$ associated to poverty dynamics f as a linear functional of the time-average of the marginal distributions of deprivation levels $\overline{f}$ and changes $h_f(u)$. We use this result to solve (DOCP).

5.1 Dominance with Binary Deprivation Levels

If $\delta_z = \delta_z^0$ is poverty indicator (headcount) the set of possible period-deprivation values is $D = \{0, 1\}$ and $U = \{-1, 0, 1\}$. From Theorem 1, since $m(1) = 1$, $m(0) = 0$ and $\phi(0) = 0$, the individual multi-period measure $M(m, \phi)$ is entirely determined by two parameters: $\phi_L \equiv \phi(1)$, the value of an increase in deprivation (loss), and $\phi_G \equiv \phi(-1)$ -the value of a decrease in deprivation (gain). In this case, as shown in the Appendix, (DOCP) is a linear program with two variables ϕ_L and ϕ_G. This variables are constrained by the monotonicity conditions (M1)-(M3) and the (LA) property, which are linear. These constraints define polytope with a set extreme points E_T given by

$$E_T = \begin{cases} \{(1,1), (-1,-1), (1,-1)\} & \text{if } T = 2 \\ \{(1/2, -1/2), (0,-1), (1,0), (1/2, 1/2)\} & \text{if } T \geq 3. \end{cases}$$

The following Theorem follows:

Theorem 2 *Suppose $\delta_z = \delta_z^0$ so that $D = \{0, 1\}$ (binary deprivation). Then poverty dynamics f^A dominates poverty dynamics f^B if and only if*

$$\overline{f}^A(1) + \phi_L h_{f^A}(1) + \phi_G h_{f_B}(-1) \leq \overline{f}^B(1) + \phi_L h_{f^B}(1) + \phi_G h_{f^B}(-1)$$

for each $(\phi_L, \phi_G) \in E_T$.

Fix a poverty dynamics f and observe that for the poverty indicator the time-average of poverty is $\overline{\pi} = 1 \times \overline{f}(1) + 0 \times \overline{f}(0) = \overline{f}(1)$. Similarly, the time-average of poverty creation between consecutive periods is $L = h_f(1)$, and each of these units contributes ϕ_L to societal multi-period deprivation. Similarly, the time-average poverty destruction over time is $G = h_f(-1)$ with each unit contributing ϕ_G. Thus, overall multi-period deprivation $Q = \overline{\pi} + \phi_L L + \phi_G G$, a weighted sum of average of poverty is $\overline{\pi}$, poverty creation L and poverty destruction G. The weights are constrained by the axioms so that $|\phi_L|, |\phi_G| \leq 1$ (monotonicity) and $\phi_L \geq \phi_G$ (loss aversion). Hence, the Theorem just says that the weighted sum for f^A must be smaller than the one for f^B.

The following Proposition is an immediate Corollary of the Theorem for $T = 2$, which we use in applications. We use $f(i, j)$ to denote the share of individuals having deprivation levels i and j in periods 1 and 2 respectively.

Proposition 1 *Suppose that $\delta_z = \delta_z^0$ and $T = 2$. Then f^A dominates f^B if and only if*

(i) $f^A(0,0) \geq f^B(0,0)$;
(ii) $f^A(0,0) + f^A(1,0) \geq f^B(0,0) + f^B(1,0)$;
(iii)$f^A(0,0) + f^A(1,0) + f^A(0,1) \geq f^B(0,0) + f^B(1,0) + f^B(0,1)$.

The Proposition unpacks the conditions of Theorem 1, by deriving explicit formulas for the bivariate poverty dynamics. Condition (i) requires A to have a larger share of individuals with incomes above the poverty line in both periods than B, i.e., more individuals who are *never poor*. Condition (ii) requires static poverty in A at period 2 to be smaller than in B.[27] Finally, using the fact that $\sum_{i,j} f(i,j) = 1$, (iii) can be restated as $f^A(1,1) \leq f^B(1,1)$. This means that society A must have lower share of individuals under the poverty line during both periods than B, i.e., less individuals who are *always poor*. We illustrate the main differences between the partial order on income processes just derived and the rankings generated using measures of social mobility and static poverty measures in our applications.

5.2 Dominance with a Continuum of Deprivation Levels

We now turn to the case in which δ_z is continuous so that $D = [0,1]$. For each number s let $[s]^+ = \max\{s, 0\}$. Fix two poverty dynamics f^A and f^B. Let $\Delta F^{A,B}(x) \equiv \overline{F}^A(x) - \overline{F}^B(x)$, $\Delta H^{A,B}(v) \equiv H_{f^A}(v) - H_{f^B}(v)$,

$$r_1^{A,B}(v) \equiv \begin{cases} [\Delta H^{A,B}(v)]^+ & \text{if } \Delta H^{A,B}(v) \geq \Delta H^{A,B}(-v) \\ [-\Delta H^{A,B}(-v)]^+ & \text{if } \Delta H^{A,B}(-v) \geq \Delta H^{A,B}(v) \end{cases} \tag{7}$$

and

$$r_2^{A,B}(v) \equiv \begin{cases} [-\Delta H^{A,B}(v)]^+ & \text{if } \Delta H^{A,B}(v) \geq \Delta H^{A,B}(-v) \\ [-\Delta H^{A,B}(-v)]^+ - [\Delta H^{A,B}(v)] & \text{if } \Delta H^{A,B}(-v) \geq \Delta H^{A,B}(v). \end{cases} \tag{8}$$

The following Theorem characterizes dominance for $T = 2$.

Theorem 3 *Fix δ_z and suppose that $D = [0,1]$ (continuum). If $T = 2$ Then f^A dominates f^B if and only if*

[27] Note that the headcount of non-poor in period 2 is precisely $f(0,0) + f(1,0)$.

(i) $\overline{F}^A(u) \geq \overline{F}^B(u)$ *(or* $\Delta F^{A,B}(u) \geq 0$*) for all* $u \in [0,1]$;
(ii) $\int_0^u \left(\Delta F^{A,B}(x) - r_1^{A,B}(x)\right) dx \geq 0$ *for all* $u \in [0,1]$;
(iii) $\int_u^1 \left(\Delta F^{A,B}(x) - r_2^{A,B}(x)\right) dx \geq 0$ *for all* $u \in [0,1]$.
If $T \geq 3$ *conditions (i)-(iii) are sufficient for dominance.*

Condition (i) says that the (time-average) cumulative distribution of deprivation levels $\overline{F}^B$ FOSD $\overline{F}^A$. This is restriction that follows from monotonicity alone. Note that the time-average of poverty for society A is $\overline{\pi}^A = 1 - \int_0^1 \overline{F}^A(u)$ and, similarly, for B we have $\overline{\pi}^B = 1 - \int_0^1 \overline{F}^B(u)$. Hence, a necessary condition implied by (i) is that $\overline{\pi}^A \leq \overline{\pi}^B$. To interpret conditions (ii) and (iii) observe that the integrands have two terms. The term $\Delta F^{A,B}(x)$ measures the difference in average cumulative deprivation between A and B. On the other hand, both $r_1^{A,B}(v)$ and $r_2^{A,B}(v)$ measure differences in the cumulative distribution of deprivation changes.

We provide some insight on the proof. In contrast to the binary case, with a continuum of deprivation levels, (DOCP) is a calculus of variations problem. In the Appendix we characterize its solution for the case of $T = 2$ in two steps.[28] First, for a fixed increasing function $m : D \to [0,1]$ with $m(0) = 0$ and $m(1) = 1$, let $\Phi(m)$ be the set of "flow-value" functions" ϕ such that the multi-period deprivation function $M = M(m, \phi)$ satisfies our axioms for any $\phi \in \Phi(m)$. We first find

$$K[m](f^A, f^B) \equiv \sup_{\phi \in \Phi(m)} K[M(m,\phi)](f^A, f^B). \qquad \text{(DOCPm)}$$

Next, given $K[m](f^A, f^B)$ for each m, we maximize across m: $K^{\max}(f^A, f^B) = \sup_m K[m](f^A, f^B)$. The first step exploits the fact that, by the monotonicity of M, m and ϕ are differentiable a.e. in their respective domains. Using a differential version of the monotonicity constraints (M1)-(M2) and the (LA) property we obtain an explicit solution of (DOCPm). To solve the second stage we use an approximation argument that relies on the fact that step functions are dense in the space of cumulative distribution functions, so that any m can be approximated by a sequence of step functions. In particular, for any fixed finite grid I^n on $[0,1]$ we define the space of cumulative distribution functions that are constant on each of the n subintervals of the grid.

[28] Finding a solution of (DOCP) for $T \geq 3$ is harder as it involes the additional monotonicity constraint (M3). This significantly complicates the characterization of the optimal control problem.

This is finite-dimensional space as each of these step functions is entirely determined by a finite-dimensional "density" vector that measure the "jumps" of the c.d.f. in each subinterval. We are able to find an explicit solution of $\sup_m K[m](f^A, f^B)$ and its value under the additional restriction that m be one of these step functions. Using an approximation argument and dominated convergence leads to a complete characterization of $K^{\max}(f^A, f^B)$. The conditions of the Theorem simply express that $K^{\max}(f^A, f^B) \leq 0$.

Theorem 1 provides an explicit characterization that can be numerically implemented. At the same time, it seems desirable to obtain conditions that allow for a more transparent comparison with measures derived in the social mobility literature and work for any T. For this purpose, we derive the order on poverty dynamics induced by the family of piece-wise linear multi-period deprivation measures $\mathcal{Q}^l \subset \mathcal{Q}^*$ introduced in section 2. These are two-parameter measures that satisfy Theorem 1 for $m(d) = d$ and $\phi(u) = 1/2(\lambda \max\{0, u\} + \gamma \min\{0, u\})$. Extending notation of section 2 for an arbitrary T, let

$$L = \sum_{t=1}^{T-1} \sum_{i:d_{it+1}-d_{it}>0} |d_{it+1} - d_{it}| \text{ and } G = \sum_{t=1}^{T-1} \sum_{i:d_{it+1}-d_{it}<0} |d_{i2} - d_{i1}|,$$

which correspond to the overall poverty creation and destruction over time. The next result characterizes the partial order $\succeq^l$ induced by $\mathcal{Q}^l$. As for the case of binary deprivation levels, the linearity of measures in λ and γ, implies that the corresponding (DOCP) is a linear program, as shown in the Appendix. The extreme points of the constraint set for λ and γ is $E_T^l = \{(1,1), (-1,1), (1,-1)\}$ if $T = 2$ and $E_T^l = \{(1/2, -1/2), (-1/2, 1/2), (0,1), (1,0)\}$ if $T \geq 3$ for $T = 2$:

Proposition 2 *Let f^A and f^B two poverty dynamics. Then $f^A \succeq^l f^B$ if and only if*

$$\overline{\pi}^A + \lambda L^A - \gamma G^A \leq \overline{\pi}^B + \lambda L^B - \gamma G^B$$

for each $(\lambda, \gamma) \in E_T^l$. In particular, if $T = 2$ this reduces to:
(i) $\pi_1^A + L^A \leq \pi_1^B + L^B$
(ii) $\pi_2^A \leq \pi_2^B$
(iii) $\pi_1^A - G^A \leq \pi_1^B - G^B$.

Proposition 2 is intuitive. Condition (i) compares poverty stocks π_1 in period one plus poverty creation L between periods 1 and 2. Condition (iii)

compares stocks poverty stocks π_1 in period one minus poverty destruction G between periods 1 and 2. The difference in sign between poverty destruction and creation across these two conditions reflects the fact that, in contrast to social or poverty mobility measures, our criterion puts different signs on the contribution of poverty inflows and outflows to social welfare. Finally, condition (ii) compares poverty levels in period 2.

6 Poverty Dynamics in the United States: 1980s versus 1990s

In this section we apply our methodology to compare poverty dynamics in the United States across two decades, the 1980s and the 1990s using data from the Panel Study of Income Dynamics (PSID). In this comparison, each decade is treated as a different "society" characterized by its own income process. We focus on the partial ordering derived from the loss aversion axiom. To apply our framework we need to specify a time window, an individual measure of income for this window, a poverty line z, and an individual measure of individual deprivation δ_z.

6.1 Poverty Status

We consider a time window of five calendar years, so that each decade is divided into two windows. For the 1980s "society", we first compute the average income between 1981 and 1985 for the 1983 cohort of the PSID. Hence, $t = 1$ is the window 1981-1985. For the same population, we compute average income between 1986 and 1990. Thus, $t = 2$ corresponds to 1986-1990 window. A deprivation path for an individual of the 1980s "society" is a pair of observations specifying the individual's poverty situation in each window of a decade. Similarly for the 1990s, $t = 1$ is the window 1991-1995 and $t = 2$ is 1996-2001. For this decade, we compute the average income in both time windows for the 1993 cohort of the PSID and, as before, we compute deprivation paths for each individual based on these measures.[29]

Following Jalan and Ravallion (1998) we consider an individual to be poor in society $A \in \{1980s, 1990s\}$ for the time window $t \in \{1, 2\}$ if the

[29] The second window was extended to 2001 because since 1997 the PSID is conducted every other year rather than on a yearly basis.

average income over a five-year window falls below the relevant poverty line of society A. There are three reasons to adopt this definition of poverty. First, it allows us to focus on changes in persistent rather than transient poverty.[30] Second, average income over five years closely follows consumption expenditure, which is what are ultimately interested in. Third, averaging income over a five-year period reduces measurement error in the identification of personal income.

In practice, as argued by Meyer and Sullivan [2006], empirical poverty measures are highly sensitive to the definitions implemented by the researcher. Our poverty estimates closely follow their recommendations to reduce the potential arbitrariness. First, we estimate poverty at the individual level using the household as the resource sharing unit. We account for different household sizes by applying the equivalence scale recommended by the National Academy of Sciences (NAS): (number of adults + number of children$\times P)^S$, where k stands for the child "consumption" proportion of an adult and S represents an economies of scale factor. We use the standard values $P = 0.7$ and $S = 0.7$.[31] Second, we use disposable income, which is post-tax and transfers income. Again, this is different from the pre-tax money income used to compute the official poverty line in United States. By using pre-tax income we would have excluded taxes and non-cash benefits such as the Earned Income Tax Credit (EITC) or housing subsidies.[32] Third, we compute real per-capita income over time using a corrected price index proposed by the Bureau of Labor Statistic (CPI-U-RS) instead of the widely used CPI-U.[33] Finally, we anchor the computation of the poverty line to 1980. This means that we take the poverty rate for 1980 to be the one estimated using the Current

[30] As shown by Bane and Ellwood [1986] and subsequent studies, transient poverty is significant in any cross-sectional analysis. In our sample, throughout these two decades the transient poor are between 2% and 3% of the population. Those in persistent poverty are between 7% and 10%.

[31] This formula imporves upon the official equivalence scale as it considers diminishing marginal cost over the whole range of family size, and extra children always cost less than extra adults.

[32] Using pre-tax income during this period to asses poverty dynamics is especially problematic as several new social policies during the 1990s took the form of non-cash benefits and the EITC was a central instrument of the Welfare Reform (Ellwood [1999]).

[33] Costa [2001] and Hamilton [2001] argue that CPI-U overstates real inflation by about 1.6 and 1.0 percentage points per year. By using the CPI-U-RS the bias is expected to be reduced by 0.4 percentage points per year. By taking the CPI-U instead of the corrected index we would overstate poverty systematically over time.

Population Survey (CPS), which is 13%. Then we computed a poverty line that generates the same poverty rate for our PSID 1980 sample.

We use z to designate the individual threshold after the equivalence scale is applied. The annual per-capita threshold computed for this sample is $z = 6,358$ in U.S. dollars of 1992. Using a five-year average y_{it} as our measure of income, an individual i is classified as poor if $y_{it} \leq z$. We provide a ranking of the poverty dynamics associated to the the two most commonly used the static poverty measures, the poverty indicator or headcount ratio, and the poverty gap.

6.2 Data

We use panel data on household income and composition for each individual from the PSID. The PSID is a widely used longitudinal survey based on interviews of the head of the household.[34] Members of the original sample are followed if they split from the original household. The PSID is considered to be nationally representative with the exception of the immigrant population. Before 1997 the surveys were conducted on a yearly basis and since then the survey was applied every other year.

A central concern in longitudinal analysis is the potential for non-random attrition. A first aspect of the problem is that panel data may not be representative of the national population after substantial cumulative attrition. According to Fitzgerald, Gottschalk and Moffitt [1998] this is not the case the PSID. A second dimension of attrition is also relevant for the estimation of parameters associated with income dynamics. In particular, panel data may be representative of the cross-sectional population, but if dynamic attrition during the period considered is not random, our sample estimates $\widehat{f}$ of the poverty dynamics distribution f will be biased. Thus, it is especially important in our case to asses the significance of this problem as the likelihood of attrition may well be associated with the income path of the attritor. We conducted an extensive analysis to asses the importance dynamic attrition which is reported in Appendix B. Our analysis suggests that selective attrition on observables is not significant in our application using the PSID.

In 1997 the PSID introduced important changes. The core sample was reduced by roughly 30% and a new sample of immigrants was incorporated. This new sample is dropped in our estimation because poverty dynamics in

[34] See see Hill [1992] for a detailed description of the survey.

our 1990s society is estimated following the 1993 cohort of the PSID. In the second time window of the 1990s decade, we used a standard imputation procedure to predict the poverty level for those who were pulled out of the sample.

6.3 Results: Ranking the 1980s versus the 1990s

6.3.1 Headcount Ratio

We compare poverty dynamics for both decades using the headcount ratio as the underlying static poverty measure. The following table shows a matrix with four entries, each representing the share of the population experiencing a particular deprivation path $f(i,j)$, where $i,j \in \{0,1\}$. We see that during the first half of the eighties, 11.3% of the sample is poor under the Jalan-Ravallion criterion. This is the percentage of the population having a five-year average income below the poverty line for the 1981-1985 period. More than half of them (5.8%), remained in this situation for the entire decade.

A few interesting facts arise when we compare both decades, and the results confirm Meyer and Sullivan [2006] findings. Namely, the stock of poverty has decreased over time. The data shows that during the 1990s both margins -poverty exits or entries- performed better compared to the 1980s: more people escaped poverty (6.0% versus 5.5%) and a smaller share of the population entered poverty (2.3% versus 3.8%).The notation $\widehat{f}$ stands for the sample estimate of the poverty dynamics f.

Table 2: *Poverty Dynamics in The United States (Headcount Ratio)*

Table 2a: 1980s

$\widehat{f}^{80s}(i,j)$	Non Poor 86-90 ($j=0$)	Poor 86-90 ($j=1$)	Total
Non Poor 81-85 ($i=0$)	0.849	0.038	0.887
Poor 81-85 ($i=1$)	0.055	0.058	0.113
Total	0.903	0.097	1.00

Table 2b: 1990s

$\widehat{f}^{90s}(i,j)$	Non Poor 96-01 ($j=0$)	Poor 96-01 ($j=0$)	Total
Non Poor 91-95 ($i=0$)	0.881	0.023	0.903
Poor 91-95 ($i=1$)	0.060	0.037	0.097
Total	0.941	0.059	1.00

We use these parameters to test the conditions stated in Proposition 1. The comparison between the 1980s and the 1990s is summarized in Table 3.

Table 3: *Comparing the 1980s and the 1990s using the Headcount Ratio*

Inequalities	1990s	1980s	Difference
$\widehat{f}(0,0)$	0.881	0.849	0.032
$\widehat{f}(0,0)+\widehat{f}(1,0)$	0.941	0.903	0.038
$\widehat{f}(0,0)+\widehat{f}(1,0)+\widehat{f}(0,1)$	0.964	0.942	0.022

The three conditions required to rank the 1990s over the 1980s are fulfilled for the sample estimates: (1) the share of never-poor population is higher, (2) final poverty is lower and (3) the share of the always-poor is lower. To establish whether these differences are statistically significant we test the null hypothesis

$$H_0 : f^{90s} \approx_{LA} f^{80s} \text{ versus } H_1 : f^{90s} \succeq f^{80s}$$

using the method suggested by Formby et al. [2004] and Kodde and Palm [1986] for inferences that involve joint equality and inequality restrictions associated with each hypothesis.[35] The method uses the vector of estimates

[35] For completeness, a detailed explanation of this procedure can be found in Apendix C.

in the last column of Table 3 and the covariance matrix associated with these parameters to construct a generalized Wald statistic W. This value is then compared to the critical value CV corresponding to a 1% confidence region. The expressions for the covariance matrix and the Wald statistic can be found in Appendix C. In this case, $W = 19.059$. The upper bound critical is $CV = 10.501$, which is obtained from Kodde and Palm [1986] using three degrees of freedom.[36] Since $W > CV$ we reject the null hypothesis and conclude that the inference $f^{90s} \succeq f^{80s}$ is robust.

6.3.2 Poverty Gap

We repeat the comparison using a different underlying static poverty measure, the poverty gap. Table 4a shows the poverty stock π_t for each time window within a decade, and the poverty creation and destruction levels L and G for each decade. The expressions required to rank both decades are summarized in Table 4b (Proposition 2).

Table 4: *Comparing the 1980s and the 1990s using the Poverty Gap*

Table 4a: Parameters

	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{L}$	$\widehat{G}$	$\frac{\pi_1+\pi_2}{2}$	$\widehat{L}+\widehat{G}$
1990s	0.028	0.017	0.007	0.017	.022	.024
1980s	0.028	0.025	0.011	0.013	.026	.024

Table 4b: Inequalities

	1990s	1980s	Difference
$\widehat{\pi}_2$	0.017	0.025	-0.008
$\widehat{\pi}_1 - \widehat{G}$	0.010	0.014	-0.004
$\widehat{\pi}_1 + \widehat{L}$	0.035	0.038	-0.004

Observe that although the initial stock of poverty is similar in both cases (0.028), the destruction of poverty is significantly higher G (0.017 versus 0.013) and the creation of poverty L is lower during the 1990s (0.007 versus 0.011).

From the sample estimates, the last column in Table 4b shows that the 1990s dominate the 1980s. This inference is significant at the 1% level. Thus,

[36] The distribution f has four parameters which sum to unity.

for any linear multi-period deprivation function that satisfies the monotonicity and loss aversion axioms the conclusion is the same as before, the 1990s dominate the 1980s.

6.4 Discussion: Comparison with Social Mobility Orders

There seems to be considerable agreement with respect to the dynamics of poverty in the United States during the 1990s. The Welfare Reform was associated with a steady reduction of the poverty rate (See, for example, Ellwood [2001] and Meyer [2006]) and a significant reduction of poverty among groups historically identified with perennial poverty (Blank [2002]). Relative to the 1980s, our own estimates show a significant increase poverty outflows but similar poverty inflows. Overall, since both poverty stocks, and the pattern of poverty creation and destruction is more favorable for the 1990s than 1980s, it is not surprising that our criterion produces a clear ranking, one consistent with the picture shared by most of the previous research. Having said this, as shown by our next application and in contrast to the results in this section, our ranking can have significant discrepancies with those the induced by looking at poverty reduction alone.

Since social mobility measures are the main instrument used to analyze income dynamics it is worth to compare the ranking induced by some of these measures relative to ours. It turns out that social mobility measures provide a rather different picture. Indeed, this is not surprising either as these measures aim to capture different information. Classic measures of social mobility measures attempt to capture the extent to which an income process leads to more or less "equalization of opportunities". The basic idea is that society A is better than B if initial condition are less correlated with final outcomes in society A than in B. From this perspective, social welfare is associated with what is often referred as "circulation". The social mobility rankings introduced by Atkinson [1983], Dandardoni [1993] and Prais-Shorrocks [1978] are in this spirit. A more recent approach introduced by Fields-Ok [1996] combines circulation and income growth (positive or negative) into a measure of "pure mobility".

Table 5: *Comparison with Social Mobility Measures*

Table 5a: Income Transition Matrix

Social Mobility Measure	Ranking Using Income Transition Matrix
Prais-Shorrocks	1980s $\succeq$ 1990s
Fields-Ok	1980s $\sim$ 1990s
Atkinson-Dardandoni	Not possible to rank

Table 5b: Poverty Transition Matrix

Social Mobility Measure	Ranking Using Poverty Transition Matrix[37]
Prais-Shorrocks	1980s$\preceq$1990s
Fields-Ok	1980s $\sim$ 1990s
Atkinson-Dardandoni	Not possible to rank

The orders induced by some of these social mobility measures are summarized in Table 5.[38] It is important to recall that none of these measures has a direct welfare implication. If anything, these figures suggest that measures of social mobility that are more heavily based on circulation rank the 1980s over the 1990s.[39] In general, there are three main reasons for this discrepancy between our criterion and the ranking produced by a social mobility measures. First, our measure focuses on the mobility of the poor and not on the entire distribution of income. Further, as many other social mobility measures, the Prais-Shorrocks and the Atkinson-Dandardoni are based on a transition matrix between income quintiles rather than "absolute" income thresholds. The second column of Table 5 reports rankings based on the social mobility measures that focus exclusively on transitions in or out of the poverty region, or within the poverty region. In particular, a "poverty transition" matrix is constructed to compute "poverty mobility" versions of the Prais-Shorrocks and the Atkinson-Dandardoni.[40] A version of the Fields-Ok "poverty mobility" index can be calculated by looking at income changes within the poverty

[37] See appendix A

[38] The mobility estimates consider the average income for each of the five-year windows defined previously for each decade. The Atkinson- Dardandoni criterion and the Prais-Shorrocks index were estimated using a quintile transition matrix. For each decade, the Fields-Ok is estimated by averaging across individuals the absolute differences in income between the two time windows.

[39] The Atkinson-Dardandoni criterion delivers a partial order much like ours and, in this case, it produces no ranking between the decades.

[40] An explicit formula for this matrix can be found in Appendix A.

region. It can be shown that this corresponds to the sum $L + G$ reported in table 4a. The results are not different than the ones obtained looking at transitions across the quintiles of the entire distribution. This means that the main discrepancies between our ranking and those based on social mobility measures are due to other reasons. Indeed, our criterion places positive value on mobility out of poverty but not mobility into poverty. Falling into poverty contributes to circulation and can thus be associated with higher values of the social mobility measures. Furthermore, our criterion takes into account the poverty stocks, not only poverty mobility. It is possible, for example, that an economy A that experiences a large reduction of poverty is ranked below other economy B for which poverty stable at a relatively low level, if the initial difference in poverty stock between A and B is sufficiently high. For this data set, the main reason for the discrepancy seems to be a combination between the latter two reasons.

7 Cross-National Comparisons: Germany, Great Britain and the United States

In this section we apply the framework in a cross-country context. We compare the poverty dynamics of the 1990s of three industrialized countries: Germany, Great Britain and the United States.

7.1 Poverty Status

The main elements defining poverty status are the same as in the previous application. We use disposable income as in most cross-national comparisons[41]. Our method compares societies having equivalent absolute poverty lines. In order to identify a cross-national comparable poor household we construct an absolute poverty line for Germany and Great Britain by applying Purchase Power Parity (PPP) to the 1992 poverty line of the United States.[42]

[41] Disposable income deduces taxes from money income.

[42] This is standard for cross-national poverty analysis but it need not be an ideal cross-national transformation of the real needs of the poor for each country because PPP takes into account all the goods and services produced in each country rather than those that are relevant for this specific population. Further, certain needs may have a cultural basis.

7.2 Data

Our analysis is based on longitudinal household survey data from the Cross-National Equivalent File (CNEF).[43] This data set is particularly well suited for our purpose since its variables are cross-nationally comparable. The CNEF contains panel data for five countries and we use information for three of them: Germany, Great Britain, and the United States.[44] For Germany, the CNEF is constructed based on the German Socio-Economic Panel (GSOEP). The GSOEP is a longitudinal data set that started in 1984 with a sample of 6,000 households living in the former Federal Republic of Germany (Western Germany). After reunification and starting in 1992 families from Eastern Germany were included to provide a nationally representative sample of unified Germany. To avoid data issues associated to this change we consider GSOEP data from 1992 to 2001. For Great Britain, the CNEF uses the British Household Panel Survey (BHPS). The BHPS began in autumn 1991 with a sample of 5,509 households selected based on postal code of residence. The resulting sample is representative of the households with postal code in England.[45] For the United States, the CNEF uses the PSID data as in the previous section.[46] The data from both the BHPS and the PSID spans from 1991 to 2000.

The PPP data used to convert the U.S. poverty line for Germany and Great Britain is from the World Bank. As in our previous application, the issues about the possibility of an estimation bias that resulting from attrition need to be addressed. Rendtel et al. [2004] conclude that there is no significant representative bias associated to attrition in the BHPS and the GSOEP files once the appropriate weights are taken into account. In the appendix we asses the importance of dynamic attrition for our estimates. We

[43] Currently the CNEF is administered by Cornell University.

[44] The other two countries in CNEF are Canada and Australia. Canada has a six year rotating sample design and Australian data is available only for the years 2001-2004. Thus the time span covered by this data is shorter than we need in order to compute the poverty dynamics distribution. Further, the data for Australia does not include the nineties.

[45] Additional samples of 1,500 households in each of Scotland and Wales were added to the main sample in 1999. In 2001 a sample of 2,000 households was added in Northern Ireland. This work does not include these data.

[46] In contrast to the PSID, which since 1997 the PSID has collected data every other year, GSOEP and BHPS collect information annually. Also, while PSID interviews the head of household, the other two panel surveys interview all household members aged 16 and older.

conclude that for the case of Great Britain there is no significant bias, but for Germany we find some evidence suggesting the opposite. We argue that the biases introduced by selective attrition in the GSOEP are predictable and play a minor role in our rankings.

7.3 Ranking Germany, Great Britain and the United States

7.3.1 Headcount Ratio

The following Table summarizes the sample estimates of the poverty dynamics distribution for each country when we the headcount ratio is used as the underlying poverty status.

Table 6: *Poverty Dynamics for Germany, Great Britain and the U.S. (Headcount Ratio)*

Table 6a: Germany

$\widehat{f}^{Ger}(i,j)$	Non Poor 97-01 ($j=0$)	Poor 97-01 ($j=1$)	Total
Non Poor 92-96 ($i=0$)	0.876	0.029	0.905
Poor 92-96 ($i=1$)	0.055	0.040	0.095
Total	0.931	0.069	1.000

Table 6b: Great Britain

$\widehat{f}^{G.B.}(i,j)$	Non Poor 96-00 ($j=0$)	Poor 96-00 ($j=1$)	Total
Non Poor 91-95 ($i=0$)	0.937	0.017	0.954
Poor 91-95 ($i=1$)	0.034	0.013	0.046
Total	0.971	0.029	1.000

Table 6c: United States

$\widehat{f}^{U.S.}(i,j)$	Non Poor 96-00 ($j=0$)	Poor 96-00 ($j=1$)	Total
Non Poor 91-95 ($i=0$)	0.884	0.022	0.906
Poor 91-95 ($i=1$)	0.058	0.036	0.094
Total	0.941	0.059	1.000

The level of poverty in Great Britain is lower than in both other countries at each point in time. In addition, the share of those initially poor who exit poverty is 72% compared to 59% and 62% for Germany and the U.S. respectively. Comparing Germany and the U.S. is also telling. While they present similar levels of initial poverty (9.4% versus 9.5%), the U.S. has more poverty destruction (5.8% versus 5.5%) and less poverty creation (2.2% versus 2.9%).[47]

Table 6 uses Proposition 1 to compare the poverty dynamics for each pair of countries. To establish the statistical significance of these comparisons we use the method described in the previous section. We conclude that the inferences $f^{GB} \succeq f^{US}$, $f^{GB} \succeq f^{Ger}$ and $f^{US} \succeq f^{Ger}$ hold at the 1% level of significance.[48].

Table 7: *Comparing Great Britain, Germany and the U.S. using the Headcount Ratio*

Parameters	G.B.	Ger	U.S.	G.B.−U.S.	G.B.−Ger	U.S.−Ger
$\widehat{f}(0,0)$	0.937	0.876	0.884	0.053	0.061	0.008
$\widehat{f}(0,0)+\widehat{f}(1,0)$	0.971	0.931	0.941	0.030	0.040	0.010
$\widehat{f}(0,0)+\widehat{f}(1,0)+\widehat{f}(0,1)$	0.987	0.960	0.964	0.023	0.027	0.004

In sum, for the headcount ratio, we find that during the poverty dynamics of these countries are ranked by our criterion as follows:

$$f^{GB} \succeq f^{US} \succeq f^{Ged}.$$

7.3.2 Poverty Gap

We extend the previous international comparison by using the poverty gap as the underlying static measure and the linear multi-period deprivation measure, as in section 5.3.1.

[47] The estimate of $\widehat{f}$ for United States is slightly different in Tables 2 and 3 as the poverty rate was anchored using the 1992 PPP oficial poverty line for United States international comparison as opposed to the 1980 CPS poverty rate anchor used previously.

[48] The corresponding Wald statistic for the hypothesis that Britain is preferred over the U.S. is greater than the critical value ($W = 409.23 > CV$). So is the one for "Britain is preferred over Germany ($W = 406.95 > CV$) and the hypothesis that U.S. dominates Germany ($W = 85.6 > CV$).

Table 9: *Comparing Great Britain, Germany and the U.S. using the Poverty Gap*

Table 8a: Parameters

	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{L}$	$\widehat{G}$	$\frac{\widehat{\pi}_1+\widehat{\pi}_2}{2}$	$\widehat{L}+\widehat{G}$
Germany	0.019	0.019	0.011	0.010	0.019	0.022
Great Britain	0.009	0.009	0.006	0.006	0.009	0.013
United States	0.027	0.017	0.007	0.017	0.022	0.024

Table 8b: Inequalities

	Ger	G.B.	U.S.	G.B.−U.S.	G.B.−Ger	U.S.−Ger
$\widehat{\pi}_2$	0.019	0.009	0.017	−0.008	−0.010	−0.002
$\widehat{\pi}_1 - \widehat{G}$	0.009	0.003	0.010	−0.007	−0.006	0.001
$\widehat{\pi}_1 + \widehat{L}$	0.030	0.015	0.034	−0.019	−0.015	0.004

Using the Poverty Gap as a measure of deprivation Great Britain continues to dominate the other two countries[49]. However, in this case Germany and the U.S. cannot be ranked. Put differently, the resulting ranking between Germany and the U.S. now depends on the value of the parameters λ and γ, i.e., the weights in the social welfare function of poverty inflows and outflows relative to the weight of poverty stocks. Indeed, Germany has lower a average poverty rate than the U.S., 0.019 compared to 0.022. On the other hand, the U.S. has more poverty destruction G, 0.017 compared to 0.010, and less poverty creation L, 0.007 compared to 0.011. It follows that if the parameters λ and γ are high enough, so that the importance of poverty flows relative to poverty stocks is high, the United States will rank better and, conversely, lower weights on flows relative to stocks will favor Germany over the U.S. in the comparison.

7.4 Discussion: Stocks and Flows, Losses and Gains

We use the cross-national comparisons above to illustrate some of the properties of our measures of poverty dynamics. We start by observing that, among

[49] As before, we derive the variance-covariance matrix directly from the parameters estimated in the sample for each country, $\widehat{d}_{it}$ and $\widehat{d}_{it+1}$, and a Wald test with three degrees of freedom allow us to reject the null hypothesis in favor of the alternative that Great Britain is prefered over Germany and the United States at a 1% significance level.

the three countries considered, the U.S. is the one that experienced highest poverty reduction during the decade, while Great Britain is the country with lowest poverty reduction. This holds using either the headcount ratio or the poverty gap. For example, from table n it is we see that the United States reduced poverty in 3.6 percentage points (from 9.4% to 5.8%), Germany by 2.6 percentage points (from 9.5% to 6.9%) and, Great Britain reduced poverty in only 1.7% percentage points (from 4.7% to 3.0%).

By contrast, our criterion ranks Great Britain first regardless of the poverty measure employed. The reason for he discrepancy between our ranking and the one induced by "poverty reduction" is simple. Our measures of poverty dynamics give weight to both poverty stocks and poverty flows and, the weight of flows relative to stocks is bounded by the monotonicity axiom. In particular, the average poverty rates across the two time windows considered for Great Britain are substantially lower than the ones of the other two countries. For example, using the headcount ratio (Table 6), Great Britain's average poverty was 3.9%, while the same numbers for the Germany and the United States were 8.2% and 7.6%, respectively. Similarly, using the poverty gap Great Britain's poverty is 0.09 less than half of the same number for Germany (0.019) and the United States (0.022).

A valid question is then to ask: Given the numbers obtained from our analysis, what would it take for a measure of poverty dynamics to produce rankings aligned with the "poverty reduction" ranking as opposed to our criterion? In concrete, what would it take to rank the U.S. poverty dynamics as "preferable" to Great Britain's one? It would require measures that violate two of our axioms. First, for changes in poverty to outweigh stocks, we would need to give up monotonicity. At the individual level this would imply that changes in the welfare attribute have potentially larger welfare implications than absolute levels (e.g. the trajectory of someone who exits poverty could be associated with higher well-being than the one of someone who is never poor). We obviously do not oppose to the idea that changes in consumption or income are carriers of utility. This is actually incorporated by our measures. However, it might be hard to argue that the effect of income or consumption changes on well-being outweigh the impact of absolute levels, especially at low levels of income and consumption. Indeed, while behavioral economics supports the importance of changes, there is a robust relationship between subjective well-being measures (a correlate of well-being) and income levels as traditionally assumed.

Yet, giving up monotonicity would not be enough rank the poverty dy-

namics of the United States over Great Britain's. Indeed, suppose we fully ignore poverty levels and focus exclusively on flows. From tables we see that, while poverty creation is lower in Great Britain than in the United States, poverty destruction is higher in the

United States. Thus, the ranking reversal would also require dropping the loss aversion axiom (at least partially).

Finally, we observe that the criterion derived from our measures produces a clear ranking between the U.S. and Germany when we use the headcount ratio. However, if the underlying static measure is the poverty gap, there is no clear ranking between these two countries. Observe that the poverty flows of the U.S. dominate those of Germany for either measure. From table n, we have that the share of individuals who exit poverty during the decade is 5.9% for the U.S. versus 5.5% in Germany. The share of the sample entering poverty is 2.2% in the U.S. and 2.8% in Germany. Similarly, using the poverty gap (Table n), poverty destruction is higher in the U.S. than Germany (0.017 versus 0.010), while poverty creation is lower in the U.S. than in Germany (0.011 versus 0.007). On the other hand, the ranking of poverty stocks depends on the static poverty measure used. The U.S. has lower average poverty rate than Germany using the headcount ratio (7.6% versus 8.2%) but the opposite is true using the poverty gap (0.019 for Germany versus 0.022 for the U.S.).[50] Hence, with the poverty gap, the subset of measures satisfying our axioms that weigh heavily stocks over flows rank Germany over U.S. Conversely, those that emphasize flows over stock put the U.S. on top.

8 Conclusion

We have provided a method to evaluate income dynamics focusing on poverty levels, inflows and outflows. The ranking based on our method differs from the ordering obtained in the social mobility literature. First, we focus on stocks and flows of poverty as opposed to transitions across all the possible categories in the distribution of income or well-being in society. Second, our criterion presumes a negative evaluation of downward mobility -mobility into poverty- and a positive evaluation of upward mobility -mobility out of poverty. In contrast, both downward and upward mobility contribute to higher social mobility.

[50] This suggests that a typical poor in the U.S. is poorer than a typical poor in Germany.

We used the method to compare and evaluate poverty dynamics across time and space. Using data for the United States we find that the poverty dynamics of the 1990s dominates the one of 1980s. Both decades exhibit similar inflows into poverty but the 1990s have considerably more outflows. This is in line with findings of the impact of the Welfare Reform. Interestingly, a number social mobility measures deliver the opposite ordering. If we hold the view that the conditions that shape the evolution of poverty were significantly improved by the reform, as argued by most of the literature (e.g. Blank [2002], Ellwood [2001]), this suggests that our method provides a more accurate account of poverty dynamics than existing measures.

Our second application compared poverty dynamics across three industrialized countries during the 1990s. We ranked Great Britain above the United States, followed by Germany. A tentative explanation is that, during this period, Great Britain had relatively good labor market opportunities for the poor and social programs to counteract poverty triggers (Hills and Waldfogel [2004]). Again, orderings based on social mobility yield a different ranking with the United States first, followed by Great Britain, and Germany comes in last.

We focused on measures satisfying axioms that we believe to be appealing but are certainly restrictive. We view these axioms as a natural benchmark that allows for a parsimonious characterization of a ranking over distributions of streams of welfare attributes, and facilitates comparisons with those singled out by social mobility. At the same time, principles that highlight other dimensions of income dynamics -e.g. the income growth rate of poor individuals rather than changes in deprivation levels- can offer important insights. Analyzing the robustness of the rankings produced by our measures as we vary the underlying static deprivation scale also deserves more attention. Expanding the set of applications is an important step for future research.

References

[1] Addison, T., D. Hulme, and R. Kanbur (2009). Poverty Dynamics: Interdisciplinary Perspectives, Oxford University Press.

[2] Alesina, A. and E. Glaeser (2004). Fighting Poverty in the US and Europe: A World of Difference, Oxford University Press.

[3] Atkinson, A.B. (1983). The Measurement of Economic Mobility, Social Justice and Public Policy, MIT Press, Cambridge, MA.

[4] Atkinson, A. B, and F. Bourguignon (1982). The Comparison of Multi-Dimensioned Distributions of Economic Status, *Review of Economic Studies*, vol. 49(2). 183-201.

[5] Autor, D., L.F. Katz, M.S. Kearney (2006). The Polarization of the US Labor Market, NBER Working Paper 11986.

[6] Bane, M. and D. Ellwood (1986). Slipping into and out of Poverty: The Dynamics of Spell, *Journal of Human Resources* 21,1-23.

[7] Blackburn, M. (1998). The Sensitivity of International Poverty Comparisons, *Review of Income and Wealth* 44, 449-472.

[8] Blank, R. (2002). Evaluating Welfare Reform in the United States, *The journal of Economic Literature*,40 1105-1166..

[9] Bourguignon, F. and S.R. Chakravarty (2003). The measurement of multidimensional poverty, *Journal of Economic Inequality 1*, 25-49.

[10] Bowman, D., D. Minehart, and M. Rabin (1999). Loss Aversion in a Consumption-Savings Model, *Journal of Economic Behavior and Organization* 38:155–178.

[11] Chapman, G. (2000). Preferences for Improving and Declining Sequences of Health Outcomes, *Journal of Behavioral Decision Making* 13, 203-18.

[12] Chetty, R. (2009a). Sufficient Statistics for Welfare Analysis, *Annual Review of Economics, 1: 451-488.*

[13] Chetty, R. (2009b). The Simple Economics of Salience and Taxation, NBER Working Paper 15246.

[14] Costa, D. (2001). Estimating Real Income in the United States from 1988 to 1994: Correcting CPI Bias Using Engel Curves, *Journal of Political Economy* 109: 1288-1310.

[15] Dardandoni, V. (1993). On Measuring Social Mobility, *Journal of Economic Theory* 61, 372-394.

[16] Di Tella, R., J. Haisken-De New, R. MacCullock (2006) , Happiness Adaptation to Income and to Status in an Individual Panel, NBER Working Paper 13159.

[17] Duclos, J.Y., D. Sahn, S.D. Younge (2006). Robust Multidimensional Poverty Comparisons, *Economic Journal*, Vol. 116, 943-968.

[18] Ellwood, D. (1999). The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements," JCPR Working Papers 124, Northwestern University/University of Chicago Joint Center for Poverty Research.

[19] Fields, G. and E.A. Ok (1996). The Meaning and Measurement of Income Mobility, *Journal of Economic Theory* 71, 349-377.

[20] Fitzgerald, J, P. Gottschalk, R. Moffit, (1998). An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics, *Journal of Human Resources* 33, 300-344.

[21] Formby, J.P., W.J. Smith and B. Zheng (2004). Mobility Measurement, Transition Matrices and Statistical Inference, *Journal of Econometrics*, 120, 181-205.

[22] Foster, J., J. Greer and E. Thorbecke (1984). A Class of Decomposable Poverty Measures, *Econometrica* 52, 761-766.

[23] Foster, J. (2007). A Class of Chronic Poverty Measures, Vanderbilt University, Working Paper No. 07-W01.

[24] Frank, R. and R.M. Hutchens (1993). Wages, Seniority, and the Demand for Rising Consumption Profiles, *Journal of Economic Behavior and Organization* 21, 251-76.

[25] Frederick, S., G. Loewenstein, T. O'Donoghue (2002). Time Discounting and Time Preference: A Critical Review, *Journal of Economic Literature XL*, 351-401.

[26] Gilboa, I., "Expectation and Variation in Multi-Period Decisions", *Econometrica*, Vol. 57, No. 5: 1153-1169.

[27] Gottschalk, P. and E. Spolaore (2002). On the Evaluation of Economic Mobility, *Review of Economic Studies* 69, No1, 191-208.

[28] Hamilton, B.W. 2001. Using Engel's Law to Estimate CPI Bias, *American Economic Review* 91: 619-630.

[29] Hill, M. (1992). *The Panel Study of Income Dynamics*, Beverly Hills: Sage Foundation.

[30] Hills, J. and J. Waldfogel (2004). A "Third Way " in Welfare Reform? Evidence from the Great Britain, *Journal of Policy Analysis and Managment* 23 n4: 765-788.

[31] Hsee, C.K., R.P. Abelson, P. Salovey (1991). The Relative Weighting of Position and Velocity in Satisfaction, *Psychological Science* 2, 263-66.

[32] Jallan, J. and M. Ravalion (1998). Transient Poverty in Postreform Rural China, *Journal of Comparative Economics*, 23, 338-357.

[33] Kahneman, D. and R. Sugden (2005). Experienced Utility as a Standard of Policy Evaluation, *Environmental and Resource Economics* 32, 161-181.

[34] Kahneman, D and A.B. Krueger (2006). Development in the Measurement of Subjective Well-Being, *Journal of Economic Perspectives* 20, 3-24.

[35] Kakwani, N. (1980). On a Class of Poverty Measures, *Econometrica* 48 (2): 437-446.

[36] Kanbur, R. (2003). Conceptual Challenges in Poverty and Inequality: One Development Economist's Perspective. mimeo.

[37] Kanbur, R. (2005). Pareto's Revenge, *Journal of Social and Economic Development* 7 (1): 1-11

[38] Kodde, D.A. and F.C. Palm (1986). Wald Criteria for Jointly Testing Equality and Inequality Restrictions, *Econometrica* 50, 1243-1248

[39] Köszegi, B.. and M. Rabin (2006). A Model of Reference-Dependent Preferences, *Quarterly Journal of Economics*, 121, 4.

[40] Loewenstein, G. and N. Sicherinan (1991). Do Workers Prefer Increasing Wage Profiles?, *Journal of Labor Economics* 9, 67-84.

[41] Meyer, B. and J. Sullivan (2006). Three Decades of Consumption and Income Poverty, Working Paper.

[42] Prais, S.J. (1955). Measuring Social Mobility, Journal of the Royal Statistical Society Series A, Part I, 118, 56-66.

[43] Read, D. (2004). Time discounting over the lifespan, *Organizational Behavior and Human Decision Processes* 94, 22-32.

[44] Rendel, U. (2002). Attrition in Household Panels: A Survey, CHINTEX Working Paper #4.

[45] Rendel, U., A. Behr, E. Bellgardt, J. Sisto (2004). Does Panel Attrition Disturb Comparative Analysis with the European Community Household Panel (ECHP)?

[46] Ross Jr., W. T. and I. Simonson (1991). Evaluations of Pairs of Experiences: A Preference for Happy Endings, *Journal of Behavioral Decision Making* 4, 273-282.

[47] Rozen, K. (2009). Foundations of Intrinsic Habit Formation, *Econometrica*, forthcoming.

[48] Samuelson, P. (1937). A Note on Measurement of Utility, *Review of Economic Studies* 4, 155-61.

[49] Sen, A.(1976). Poverty: An Ordinal Approach to Measurement, *Econometrica* 44, 219-231.

[50] Shalev, J. (1997) "Loss Aversion in a multi-period Model", *Journal of Mathematical Social Sciences* 22: 203-226.

[51] Shorrocks, A.F.(1978). The Measurement of Mobility, *Econometrica* 46, 1013-1024.

[52] Stein, E. M., Shakarchi, R. (2005). Real Analysis, Priceton University Press.

[53] Thorbecke, E. (2004,). Conceptual and Measurement Issues in Poverty Analysis, Discussion Paper No 2004/04, WIDER.

[54] Varey, C. A. and D. Kahneman (1992). Experiences Extended Across Time: Evaluation of Moments and Episodes, *Journal of Behavioral Decision Making* 5,169-85.

A Proofs

A.1 Representation

Theorem 1 is an immediate consequence of Lemmas 1 and 2 in the main text. We provide a proof these lemmas below. By axiom (SC), for any multi-period deprivation measure $q : Y^T \to \mathbb{R}$ there exists a static deprivation index δ_z and a monotonic function $M : \delta_z(Y)^T \to [0,1]$ such that $q(\mathbf{y}) = M(\delta_z(y_1), ..., \delta_z(y_T))$. We write $q[M, \delta_z]$ to designate a multi-period deprivation measure that satisfies these axioms for M and δ_z.

A.1.1 Proof of Lemma 1

The proof relies on two lemmas. The first one shows that measures satisfying (SF) allow for representation with two separate components, one that depends only on levels of deprivation regardless of there sequencing along a trajectory, and the other one that depends exclusively on changes in deprivation over time. The second result uses the (TD) axiom to show that measures satisfy a time-separability property.

Lemma 4 *If $q[M, \delta_z]$ satisfies axiom (SF) then there exist $S : [0,1] \to \mathbb{R}$, and $\Phi : U \to \mathbb{R}$ such that*

$$M(d_1, ..., d_T) = S\left(\nu(\mathbf{d})\right) + \Phi(d_2 - d_1, ..., d_T - d_{T-1}),$$

where $\nu : D^T \to [0,1]$ is such that $\mu(\mathbf{y}) = \nu(\delta_z(y_1), ..., \delta_z(y_T))$.

Proof. Fix M and δ_z. To establish the Lemma we start with a simple observation. Let $\mu : Y^T \to [0,1]$ be the map that satisfies axiom (SF). By the (SC) axiom we must have that $\mu(\mathbf{y}) = \nu(\delta_z(y_1), ..., \delta_z(y_T))$ for some map $\nu : D^T \to [0,1]$. Since, μ is strictly increasing in an argument y_t if $y_t < z$ then ν is strictly increasing in all of its arguments. Now, consider the map $B : D^T \to [0,1] \times U$ defined by $B(d_1, ..., d_T) = (\nu(d_1, ..., d_T), w(d_1, ..., d_T))$ where ν is the map defined above and $w(d_1, ..., d_T) = (d_2 - d_1, ..., d_T - d_{T-1})$. Since ν is strictly increasing, B is a bijection. Thus, there exists $N_M : [0,1] \times U \to [0,1]$ such that

$$M(d_1, ..., d_T) = N_M(\nu(d_1, ..., d_T), w(d_1, ..., d_T)). \tag{9}$$

for all $(d_1, ..., d_T)$. Using (9), we can rewrite the condition (SF2) of axiom (SF) in terms of ν. In particular, $q[M, \delta_z]$ satisfies $(SF2)$ if and only if for each $\nu, \nu' \in [0,1]$ and $w, \widetilde{w} \in U$ we have that

$$N_M(\nu, w) \geq N_M(\nu', w) \Longrightarrow N_M(\nu, \widetilde{w}) \geq N_M(\nu', \widetilde{w}),$$

a standard additive separability condition. It follows that there exist S and Φ such that $N_M(\nu, w) = S(\nu) + \Phi(w)$. ■

Lemma 5 *If q satisfies axiom (TD) then there exist a collection of functions $\{q_t\}_{t=1}^{T-1}$ with $q_t : Y^2 \to \mathbb{R}$ (unique up to a common affine transformation) such that $q(\mathbf{y}) = \sum_{t=1}^{T-1} q_t(y_t, y_{t+1})$ for any $y \in Y^T$.*

Proof. Let $h : Y^T \to Y^{2(T-1)}$ be the map that assigns to each income stream $\mathbf{y} = (y_1, ..., y_T)$ the vector $h(\mathbf{y}) = ((y_1, y_2), (y_2, y_3), ..., (y_{T-1}, y_T))$ of ordered pairs corresponding to consecutive time periods. Note that h is injective and its image is $H = \{\mathbf{h} = (h_1, ..., h_{T-1}) \in Y^{2(T-1)} | h_t = (h_{t,1}, h_{t,2}),\ h_{t,2} = h_{t+1,1}\}$. In particular, $h : Y^T \to H$ is a bijection. Denote by $\succcurlyeq$ the ranking on Y^T associated to $q : \mathbf{y}' \succcurlyeq \mathbf{y}$ iff $q(\mathbf{y}) \geq q(\mathbf{y}')$. This order induces an order on H: $h(\mathbf{y}) \succcurlyeq_H h(\mathbf{y}')$ iff $\mathbf{y}' \succcurlyeq \mathbf{y}$. We establish the lemma by showing that axiom (TD) induces a standard additive separability condition on $\succcurlyeq_H$.

Indeed, the (TD) axiom on $\succcurlyeq$ implies that $\succcurlyeq_H$ satisfies the following property: For any non-empty time interval A and $\mathbf{h}, \mathbf{l}, \mathbf{k} \in H$ such that we have that

$$(\mathbf{k}_A, \mathbf{h}_{\sim A}) \succcurlyeq_H (\mathbf{k}_A, \mathbf{l}_{\sim A}) \Leftrightarrow (\mathbf{k}'_A, \mathbf{h}_{\sim A}) \succcurlyeq_H (\mathbf{k}'_A, \mathbf{l}_{\sim A})$$

for any $\mathbf{k}'_A \in H$. It follows from Arrow (1959) that there exist a collection of functions $\{q_t\}_{t=1}^{T-1}$ (unique up to a common affine transformation) such that if $\mathbf{h} = (h_1, ..., h_{T-1})$ then

$$q_H(\mathbf{h}) = \sum_{t=1}^{T-1} q_t(h_t)$$

represents $\succcurlyeq_H$. Since h is a bijection, we conclude that

$$q(\mathbf{y}) \equiv q_H(h(\mathbf{y})) = \sum_{t=1}^{T-1} q_t(y_t, y_{t+1})$$

represents $\succcurlyeq$. ■

Proof of Lemma 1.

We use the notation introduced in the proof of Lemma 4. Let $q[M, \delta_z]$ be a multi-period deprivation measure that satisfies axioms (SC) and (TD). From Lemma 4, there exist maps S and Φ such that $M(\mathbf{d}) = S(\nu(\mathbf{d})) + \Phi(w(\mathbf{d}))$, where $\nu(\cdot)$ is a strictly increasing map that depends exclusively on the distribution of deprivation levels associated to $\mathbf{d}$ (in particular, it is invariant to permutations of the components of vector $\mathbf{d}$) and $w(\mathbf{d}) = (d_2 - d_1, ..., d_T - d_{T-1})$. Combining this with Lemma 5 it follows that there exist a collection of functions $\{s_t\}_{t=1}^{T-1}$ and a collection $\{\phi_t\}_{t=1}^{T-1}$ such that that $S(\nu(\mathbf{d})) = \sum_{t=1}^{T-1} s_t(d_t, d_{t+1})$ and $\Phi(w(\mathbf{d})) = \sum_{t=1}^{T-1} \phi_t(d_{t+1} - d_t)$.

To establish the lemma we show that $s_t(d_t, d_{t+1})$ is additively separable. Indeed, if the latter holds then $S(\nu(\mathbf{d})) = \sum_{t=1}^{T} \widetilde{m}_t(d_t)$ for a collection of functions $\{s_t\}_{t=1}^{T-1}$. Now, since $S(\nu(\mathbf{d}))$ depends exclusively on the distribution of deprivation levels associated to $\mathbf{d}$, we must have that $\widetilde{m}_t = \widetilde{m}$, i.e., it is independent of t.

We first show that s_1 and s_{T-1} -the "extremes", must be separable and then conclude by induction. Consider the path $\mathbf{d} = (d, d_2, ...d_{T-1}, d')$ and $\mathbf{d}_\pi = (d', d_2, ...d_{T-1}, d)$. Thus, $\mathbf{d}_\pi$ is identical to $\mathbf{d}$ except for the fact that the first and last component have been permuted. Since, $\nu(\cdot)$ is invariant to permutations we must have that $S(\nu(\mathbf{d})) = S(\nu(\mathbf{d}_\pi))$, which translates to $s_1(d, d_2) - s_1(d', d_2) = s_{T-1}(d_{T-1}, d) - s_{T-1}(d_{T-1}, d')$. Since this equality holds for any d and d' regardless of the values of d_2 and d_T, we conclude that the right hand side is independent of d_2 and the left hand side is independent of d_{T-1}. It follows that s_1 and s_{T-1} are additively separable. Suppose that s_t

is additively separable for all $t \leq k-1$ with $k > 1$. That is, $s_t(d, d') = a_t(d) + b_t(d')$. We claim that s_k is additively separable. Towards that end consider an arbitrary path $\mathbf{d} = (d_1, ..., d_T)$ and $\mathbf{d}_\pi^k$ which is identical to $\mathbf{d}$ except for the fact that the the first and the kth component have been permuted. As before, we must have that $S(\nu(\mathbf{d})) = S(\nu(\mathbf{d}_\pi^k))$ from which, using the induction hypothesis we get $s_k(d_1, d_{k+1}) - s_k(d_k, d_{k+1}) = -a_1(d_k) + a_1(d_1) + b_{k-1}(d_k) - b_{k-1}(d_1)$. Thus, $s_k(d_1, d_{k+1}) - s_k(d_k, d_{k+1})$ is independent of d_{k+1} and this holds for any d_1 and d_k. It follows that s_k is additively separable as desired. We conclude that $M(\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} m(d_t) + \sum_{t=1}^{T-1} \phi_t(d_{t+1} - d_t)$ for some monotonic function m and a collection of functions $\{\phi_t\}_{t=1}^{T-1}$.

Finally, observe that the representation is unique up to an affine a common affine transformation for m and the ϕ_t functions, which allows us to impose the normalization $M(1, ..., 1) = m(1) + \sum_{t=2}^{T} \phi_t(0) = 1$ and $M(0, ..., 0) = m(0) + \sum_{t=2}^{T} \phi_t(0) = 0$. It follows that $m(1) - m(0) = 1$. Without loss of generality we can always normalize $m(0) = \phi_t(0) = 0$ (as $M(\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} m(d_t) - m(0) + \sum_{t=1}^{T-1} \phi_t(d_{t+1} - d_t) - \phi_t(0)$). ■

A.1.2 Proof of Lemma 2

We establish the lemma for the case $\delta_z(Y) = [0, 1]$. The case in which $\delta_z(Y) = \{0, 1\}$.(poverty indicator) is simpler and omitted.

Lemma 6 *If $q[M, \delta_z]$ satisfies axioms (TD), (SF), and (CN) then the representation (5) must have $\phi_1 = \phi_{T-1}$.*

Proof. We provide a detailed argument for T odd. The argument for T even is very similar and, thus, omitted. The result is established by comparing formula (5) for a family of paths that, given axiom (CN), are associated with the same multi-period deprivation. Indeed, fix any $\epsilon > 0$ such that $(T-1)\epsilon \leq 1$. We consider two paths $\mathbf{d}^\epsilon$ and $\widehat{\mathbf{d}}^\epsilon$ that induce the same distributions over deprivation levels and over deprivation differences . The path $\mathbf{d}^\epsilon$ is defined by $d_1^\epsilon = (T-1)\epsilon$, $d_2^\epsilon = 0$, and $d_{k+1}^\epsilon = d_k^\epsilon + \epsilon$ for $k \geq 2$. That is, $\mathbf{d}^\epsilon$ described a sharp decrease of $(T-1)\epsilon$ in deprivation from period 1 to period 2 and a constant increase of ϵ in per-period- deprivation thereafter. The path $\widehat{\mathbf{d}}^\epsilon$ is given by $\widehat{d}_1^\epsilon = \epsilon$, $\widehat{d}_k^\epsilon = \widehat{d}_{k-1}^\epsilon + \epsilon$ for $k \in \{1, ..., T-1\}$, and $\widehat{d}_T^\epsilon = 0$. Note that $\widehat{d}_{T-1}^\epsilon = (T-1)\epsilon$, so $\widehat{\mathbf{d}}^\epsilon$ describes a path that increases ϵ each period from $t = 1$ to $t = T-1$ and falls sharply in the last period. By construction, both

paths have the same "support" of deprivation levels $\{0, \epsilon, ..., (T-1)\epsilon\}$, and deprivation differences satisfy $d_2^\epsilon - d_1^\epsilon = \widehat{d}_T^\epsilon - \widehat{d}_{T-1}^\epsilon = -(T-1)\epsilon$, $d_T^\epsilon - d_{T-1}^\epsilon = \widehat{d}_2^\epsilon - \widehat{d}_1^\epsilon = \epsilon$, and $d_t^\epsilon - d_{t-1}^\epsilon = \widehat{d}_t^\epsilon - \widehat{d}_{t-1}^\epsilon$ for $t \in \{3, ..., T-1\}$. By the (CN) axiom, since $\mathbf{d}^\epsilon$ and $\widehat{\mathbf{d}}^\epsilon$ induce the same distribution over deprivation levels and over deprivation differences, $M(\mathbf{d}^\epsilon) = M(\widehat{\mathbf{d}}^\epsilon)$. Using (5) and letting $x = (T-1)\epsilon$, this translates into

$$\phi_1(-x) - \phi_{T-1}(-x) = \phi_{T-1}\left(\frac{x}{T-1}\right) - \phi_1\left(\frac{x}{T-1}\right). \tag{10}$$

Since ϵ is any arbitrary positive number such that $(T-1)\epsilon \leq 1$, 10 holds for all $x \in (0, 1]$.

A similar equation obtains if we consider two different paths $\mathbf{a}^\epsilon$ and $\widehat{\mathbf{a}}^\epsilon$ that also induce the same distributions over levels and differences. Indeed, let $a_1^\epsilon = 0$, $a_2^\epsilon = (T-1)\epsilon$, and $a_k^\epsilon = a_{k-1}^\epsilon - \epsilon$ ($\mathbf{a}^\epsilon$ describes a sharp increase from period 1 to period 2 and steady decrease thereafter); and $\widehat{a}_1^\epsilon = (T-2)\epsilon$, $\widehat{a}_k^\epsilon = \widehat{a}_{k-1}^\epsilon - \epsilon$ for $k \in \{2, ..., T-1\}$, and $\widehat{a}_T^\epsilon = (T-1)\epsilon$ ($\widehat{\mathbf{a}}^\epsilon$ describes a steady decrease from period 1 to period $T-1$ and sharp increase at T). By axiom (CN), since $M(\mathbf{a}^\epsilon) = M(\widehat{\mathbf{a}}^\epsilon)$, using (5) we get

$$\phi_1(x') - \phi_{T-1}(x') = \phi_1\left(\frac{-x'}{T-1}\right) - \phi_{T-1}\left(\frac{-x'}{T-1}\right) \tag{11}$$

for any $x' \in (0, 1]$. Combining (10) with (11) for $x' = \frac{x}{T-1}$ we get $\phi_1(-x) - \phi_{T-1}(-x) = \phi_1\left(\frac{-x}{(T-1)^2}\right) - \phi_{T-1}\left(\frac{-x}{(T-1)^2}\right)$. Iterating this formula n times yields

$$\phi_1(-x) - \phi_{T-1}(-x) = \phi_1\left(\frac{-x}{(T-1)^{2n}}\right) - \phi_{T-1}\left(\frac{-x}{(T-1)^{2n}}\right).$$

Letting $n \to \infty$ and invoking the continuity of M (and thus ϕ_1 and ϕ_{T-1}) gives $\phi_1(-x) - \phi_{T-1}(-x) = \phi_1(0) - \phi_{T-1}(0) = 0$. Similarly, combining (11) with (10) for $x = \frac{x'}{T-1}$ and repeating the same argument above gives $\phi_1(x') - \phi_{T-1}(x') = 0$. In sum, $\phi_1(x) = \phi_{T-1}(x)$ for all $x \in [-1, 1]$ as desired. ■

Proof of Lemma 2. Fix $x \in (0, 1]$. To show that $\phi_t = \phi$ for all $t \in \{1, ..., T-1\}$ it is sufficient to establish that for each $x > 0$ that there exists two numbers b_x and c_x such that $\phi_t(x) = b_x$ and $\phi_t(-x) = c_x$ for all t.

Since x is fixed throughout the proof, to ease notation, let $b_t \equiv \phi_t(x)$ and $c_t \equiv \phi_t(-x)$. Our aim is to show that the set of $2(T-1)$ variables defined by the b_t's and c_t's is determined by two free parameters.

We start by invoking lemma 6, from which

$$b_1 = b_{T-1} \text{ and } c_1 = c_{T-1}. \tag{12}$$

Now, consider two sets of (individual) deprivation paths Π_I and Π_R . The first set, Π_I, consists of $(T-2)$ "poverty-increase blip" paths, indexed by $j \in \{2, ..., T-1\}$. Each path $\mathbf{d}^j \in \Pi_I$ is defined by $d_t = 0$ for all $t \neq j$ and $d_t = x$ for $t = j$, i.e. a path that enters poverty reaching a deprivation level x at time $t = j$ and exits poverty thereafter. By construction, each of these paths induces the same distribution on the set of deprivation levels (mass $1/T$ on x and $1-1/T$ on 0) and the same distribution on deprivation changes (mass $1/T-1$ on $-x$ and x respectively, and $1-2/T-1$ on 0). Thus, by the (CN) axiom any path in Π_I has the same deprivation level. We can summarize this by writing $M(\mathbf{d}^j) = M(\mathbf{d}^{j+1})$ for all $\mathbf{d}^j \in \Pi_I$, $j = \{2, ..., T-2\}$. Using the formula 5 the previous yields a system of $(T-3)$ linearly independent equations of the form

$$b_j + c_{j+1} = b_{j+1} + c_{j+2}. \tag{13}$$

The second set of paths is Π_R, which consists of $(T-2)$ "poverty-reduction blip" paths, also indexed by $j \in \{2, ..., T-1\}$. Each path $\mathbf{d}^j \in \Pi_R$ is defined by $d_t = 1$ for all $t \neq j$ and $d_t = 1-x$ for $t = j$, i.e. a path associated with a single deprivation reduction of magnitude x at time $t = j$. As before, by construction, all paths in Π_R induce the same distribution on deprivation levels and changes. It follows that $M(\mathbf{d}^j) = M(\mathbf{d}^{j+1})$ for all $\mathbf{d}^j \in \Pi_R$, $j = \{2, ..., T-2\}$. Using formula 5 we now obtain the set of $(T-3)$ linearly independent equations

$$c_j + b_{j+1} = c_{j+1} + b_{j+2}. \tag{14}$$

The system defined by (12), (13) and (14) defines $2(T-2)$ linearly independent equations. Thus, there are at most 2 degrees of freedom in the $2(T-1)$ variables defined by the b's and the c's. Further, it is immediate that $b_j = b_x$ for all j and $c_j = c_x$ for all j is two-parameter solution of this system for any values of b_x and c_x. The proof is complete. ■

A.2 multi-period Poverty Orders

Lemma 3 *Suppose that M satisfies (6) for the pair (m, ϕ). Then*

$$Q[M](f) = \int_0^1 m(x)\overline{f}(x)dx + \int_{-1}^1 \phi(v)h_f(v)dv.$$

Proof. If M satisfies (6) then

$$Q[M](f) = \frac{1}{T}\sum_{t=1}^{T}\int_{D^T} m(x_t)f(\mathbf{x})d\mathbf{x} + \frac{1}{T}\sum_{t=1}^{T-1}\int_{D^T} \phi(x_{t+1} - x_t)f(\mathbf{x})d\mathbf{x}.$$

Integrating over all periods other than t, we get $\int_{D^T} f(\mathbf{x})m(x_t)d\mathbf{x} = \int_D f_t(x)m(x)dx$. Similarly, integrating over all periods other than t and $t-1$ we get

$$\int_{D^T} f(\mathbf{x})\phi(x_{t+1}-x_t)d\mathbf{x} = \int_{D^2} f_{t,t+1}(x,x)\phi(x'-x)dxdx' = \int_U \phi(v)h_{t,t+1}[f](v)dv.$$

Combining the previous,

$$\begin{aligned} Q[M](f) &= \frac{1}{T}\sum_{t=1}^{T}\int_D f_t(x)m(x)dx + \frac{1}{T}\sum_{t=1}^{T}\int_U \phi(v)h_{t,t+1}[f](v)dv \\ &= \int_D \left[m(x)\frac{1}{T}\sum_{t=1}^{T} f_t(x)\right]dx + \int_U \left[\phi(v)\frac{1}{T}\sum_{t=1}^{T-1} h_{t,t+1}[f](v)\right]dv \\ &= \int_D m(x)\overline{f}(x)dx + \int_U \phi(v)h_f(v)dv, \end{aligned}$$

as desired. ■

Recall that, given a poverty dynamics f, $H_f(v) = \int_{-1}^{v} h_f(u)du$.

A.3 Proof of Theorem 2

Theorem 2. *Suppose that $D = \{0, 1\}$ (binary deprivation). Then poverty dynamics f^A dominates poverty dynamics f^B if and only if*

$$\overline{f}^A(1) + \phi_L h_{f^A}(1) + \phi_G h_{f_B}(-1) \leq \overline{f}^B(1) + \phi_L h_{f^B}(1) + \phi_G h_{f^B}(-1)$$

for each $(\phi_L, \phi_G) \in E_T$.

Proof. If $D = \{0, 1\}$ then the set of deprivation changes is $U = \{-1, 0, 1\}$. Thus a deprivation function M is entirely determined by two parameters, $\phi(1)$ and $\phi(-1)$. Indeed, in this case m is entirely determined by the normalization $m(0) = 0$ and $m(1) = 1$. Note also that since $\phi(0) = 0$, $\phi(1)$ and $\phi(-1)$ are the only free parameters. In this case, $K[M](f^A, f^B) = \Delta\overline{f} + \phi(1)\Delta h(1) + \phi(-1)\Delta h(-1)$, where $\Delta\overline{f} = \overline{f}^A(1) - \overline{f}^B(1)$ and $\Delta h(v) = h_{f^A}(v) - h_{f^B}(v)$ for $v \in \{-1, 1\}$.

To characterize the constraints on $\phi(1)$ and $\phi(-1)$, observe that in this case the (LA) property translates into $\phi(1) \geq \phi(-1)$ (poverty inflows yields more deprivation than a poverty outflow). We just need to check the monotonicity restrictions for $d' = 1$ and $d = 0$ for $d_+, d_- \in \{0, 1\}$. Conditions (M1) and (M2) yield the same constraints: (M1') $\phi(1) \leq 1$ and (M2') $\phi(-1) \geq -1$. Condition (M3) yields two additional constraints: (M3') $|\phi(1) - \phi(-1)| \leq 1$ and (M3") $|\phi(1) + \phi(-1)| \leq 1$. Note that

From the previous, optimization problem (DOCP) becomes

$$K(f^A, f^B) \equiv \max_{\phi(1),\phi(-1)} \Delta\overline{f} + \phi(1)\Delta h(1) + \phi(-1)\Delta h(-1)$$

subject to

$$\begin{array}{ll} \text{(LA)} & \phi(1) \geq \phi(-1) \\ \text{(M1')} & \phi(1) \leq 1 \\ \text{(M2')} & \phi(-1) \geq -1 \end{array}$$

and, if $T \geq 3$

$$\begin{array}{ll} \text{(M3')} & |\phi(1) - \phi(-1)| \leq 1 \\ \text{(M3")} & |\phi(1) + \phi(-1)| \leq 1. \end{array}$$

This is a linear program and the set E_T is the simply set of extreme points of the polytope defined by the linear constraints above (The set is different for $T = 2$ and $T \geq 3$ as constraints (M3') and (M3") apply only to $T \geq 3$.) Since the solution the linear program is an extreme point of this set, the result follows. ■

A.4 Proof of Theorem 3

The Theorem follows from the following Lemmata:

Lemma 7 *Let $\theta(v) \equiv \phi(v) - \phi(-v)$. Suppose that $m\phi$ and ϕ are differentiable almost everywhere. The pair (m, ϕ) satisfies monotonicity constraints*

(M1)-(M2) and ϕ satisfies the (LA) property if and only if (LA) $\theta'(v) \geq 0$ and the following conditions are satisfied:

(M1') $\theta'(v) - \phi'(-v) \leq a_m(v) \equiv \min_{x\in[0,1-v]} m'(x)$;

(M2') $\phi'(-v) \leq \alpha_m(v) \equiv \min_{x\in[v,1]} m'(x)$.

Proof. It is immediate that the (LA) property can be restated in differential terms as $\theta'(v) \geq 0$ for all $v \geq 0$. Observe that $\theta'(v) = \phi'(v) + \phi'(-v)$.

A differential version of the monotonicity condition (M1) is $\phi'(d^+ - d) \leq m'(d)$ for all $d, d^+ \in [0,1]$. This can be broken down into two parts: for each $u \geq 0$ we have that (M1') $\phi'(u) \leq m'(x)$ for each $x \in [0, 1-u]$, and (M2') $\phi'(-u) \leq m'(x)$ for each $x \in [u, 1]$. Equivalently, (M1') $\theta'(v) - \phi'(-v) \leq a_m(v) \equiv \min_{x\in[0,1-v]} m'(x)$; and (M2') $\phi'(-v) \leq \alpha_m(v) \equiv \min_{x\in[v,1]} m'(x)$.

Now, condition (M2) is $-\phi'(d^+ - d) \leq m'(d^+)$ for all $d, d^+ \in [0,1]$. This condition can be expressed as: for each $u \geq 0$, (i) $-\phi'(u) \leq m'(x)$ for $x \in [u,1]$ and (ii) $-\phi'(-u) \leq m'(x)$ for $x \in [0, 1-u]$. Observing that $-\phi'(-u) = -\theta'(u) + \phi(u)$, condition (ii) is implied by (M1') and (LA). Similarly, condition (i) is implied by (M2') and (LA). ■

Lemma 8 *Let $r_1(x)$ and $r_2(x)$ be the maps defined in the main text. If $T = 2$ the value of (DOCPm) is given by*

$$K[m](f^A, f^B) = \int_0^1 m(x)\Delta\overline{f}(x)dx + \int_0^1 \left[\alpha_m(x)r_1(x) + a_m(x)r_2(x)\right] dx$$

where $a_m(x) \equiv \min_{v\in[0,1-x]} m'(v)$ and $\alpha_m(x) = \min_{v\in[x,1]} m'(v)$.

Proof. Using Lemma 3 we have that

$$K[M(m,\phi)](f^A, f^B) = \int_0^1 m(x)\Delta\overline{f}(x)dx + \int_{-1}^1 \phi(v)\Delta h(v)dv, \tag{15}$$

where and $\Delta h(v) = h_{f^A}(v) - h_{f^B}(v)$. Since f^A and f^B are fixed, throughout the rest of the proof we omit the dependency of Δh and ΔH on f^A and f^B for ease of notation. Recall that $K[m](f^A, f^B) \equiv \sup_{\phi\in\Phi(m)} K[M(m,\phi)](f^A, f^B)$, where $\Phi(m)$ is set of flow functions such that ϕ and m are consistent with Theorem 1. That is, $\phi \in \Phi(m)$ if satisfies the (LA) property and the pair (m, ϕ)

satisfies the monotonicity restrictions (M1)-(M2). Letting $K_\phi[m](f^A, f^B) \equiv \sup_{\phi\in\Phi(m)} \int_{-1}^{1} \phi(v)\Delta h(v)dv$, from (15) we get

$$K[M(m,\phi)](f^A, f^B) = \int_0^1 m(x)\Delta\overline{f}(x)dx + K_\phi[m](f^A, f^B). \qquad (16)$$

Now, for $D = [0,1]$ any deprivation function M is continuous and increasing in each of its arguments. Thus, it is Frechet-differentiable almost everywhere (Lebesgue) on D. It follows that if $M = M(m,\phi)$, m and ϕ are differentiable a.e. as well.

Let $\Delta H(v) = H_{f^A}(v) - H_{f^B}(v)$. Integration by parts, yields

$$\int_{-1}^{1} \phi(v)\Delta h(v)dv = -\phi(-1)\Delta H(-1) - \int_{-1}^{1} \phi'(v)\Delta H(v)dv$$

where we used $\Delta H(1) = H_{f^A}(1) - H_{f^B}(1) = 0$ (as $H_f(1) = 1$) and $\Delta H(-1) = 0$. Using $\theta(v) \equiv \phi(v) - \phi(-v)$,

$$\begin{aligned}
\int_{-1}^{1} \phi'(v)\Delta H(v)dv &= \int_0^1 \phi'(-v)\Delta H(-v)dv + \int_0^1 \phi'(v)\Delta H(v)dv \\
&= \int_0^1 \phi'(-v)\Delta H(-v)dv + \int_0^1 \left[\theta'(v) - \phi'(-v)\right]\Delta H(v)dv \\
&= -\int_0^1 \phi'(-v)\Delta\widehat{H}(v)dv + \int_0^1 \theta'(v)\Delta H(v)dv
\end{aligned}$$

where $\Delta\widehat{H}(v) \equiv \Delta H(v) - \Delta H(-v) = \int_{-v}^{v} \Delta h(u)du$.

As shown by Lemma 7, the constraints (LA), (M1') and (M2') depend on θ' and ϕ' alone and so does the above expression. Let $\gamma_1(v) \equiv \theta'(v)$ and $\gamma_2(v) \equiv \phi'(-v)$. Combining the previous,

$$K_\phi[m](f^A, f^B) = \sup_{\gamma_1,\gamma_2} -\int_0^1 \gamma_1(v)\Delta H(v)dv + \int_0^1 \gamma_2(v)\Delta\widehat{H}(v)dv$$

subject to

(LA)	$\gamma_1(v) \geq 0$ for all $v \geq 0$
(M1')	$\gamma_1(v) - \gamma_2(v) \leq a_m(v)$ for all $v \geq 0$
(M2')	$\gamma_2(v) \leq \alpha_m(v)$ for all $v \geq 0$.

The solution of this program (which is linear in γ_1 and γ_2) has $\gamma_2^*(v) = \alpha_m(v)$ for $\Delta\widehat{H}(v) \geq 0$ and $\gamma_2^*(v) = \gamma_1^*(v) - a_m(v)$ if $\Delta\widehat{H}(v) \leq 0$. Letting $\chi(v) = 1$ if $\Delta\widehat{H}(v) \geq 0$ and $\chi(v) = 0$ otherwise, the previous translates into

$$\gamma_2^* = \chi\alpha_m + (1-\chi)(\gamma_1^* - a_m). \tag{17}$$

Using this to eliminate γ_2^* in the objective, after some manipulation it follows that

$$\gamma_1^* = \varepsilon(a_m + \alpha_m), \tag{18}$$

where $\varepsilon(v) = 1$ if $(1-\chi(v))\Delta\widehat{H}(v) - \Delta H(v) \geq 0$, and $\varepsilon(v) = 0$ otherwise. Using the solution (17) and (18) to evaluate $K_\phi[m](f^A, f^B)$, we find that the integrand that is linear in both a_m and α_m. After some manipulation we get

$$K_\phi[m](f^A, f^B) = \int_0^1 \left[\alpha_m(x)r_1(x) + a_m(x)r_2(x)\right] dx$$

as desired. ■

Lemma 9 *Let* $\Delta F^{A,B}(x) \equiv \overline{F}^A - \overline{F}^B$, $l_1^{A,B}(u) \equiv \max_{u\in[0,1]} \frac{1}{1-u}\int_0^{1-u}\left(r_1^{A,B}(x) - \Delta F^{A,B}(x)\right) dx$ *and* $l_2^{A,B}(u) \equiv \max_{u\in[0,1]} \frac{1}{1-u}\int_u^{1}\left(r_2^{A,B}(x) - \Delta F^{A,B}(x)\right) dx$. *Then*

$$K(f^A, f^B) = \max_{u\in[0,1]}\{l_1^{A,B}(u), l_2^{A,B}(u), -\Delta F^{A,B}(u)\}.$$

Proof. The proof has three steps. Step 1 uses Lemma 8 to show that $K(f^A, f^B)$ can be derived as the value of a calculus of variations problem on the set of distributions on $[0,1]$, which is denoted $\Delta_{[0,1]}$. Steps 2 and 3 go on to show that the value of this problem is bounded above and below by $\max_{u\in[0,1]}\{l_1(u), l_2(u), -s(u)\}$.For ease of notation throughout the proof we use $s \equiv \overline{F}^A - \overline{F}^B$ and note that $s(1) = 0$ as $\overline{F}(1) = 1$ for any poverty dynamics f.

For each p.d.f. $\lambda \in \Delta_{[0,1]}$ consider the functional $C(r, s; \lambda) \equiv \int_0^1 (\alpha_\lambda(x)r_1(x) + a_\lambda(x)r_2(x))dx - \lambda$ with $\alpha_\lambda(x) = \min_{u\in[0,1-x]} \lambda(u)$ and $a_\lambda(x) = \min_{u\in[x,1]} \lambda(u)$.. We observe that since $\alpha_\lambda(x), a_\lambda(x) \leq \lambda(x)$, both of these maps are uniformly bounded by $\int_0^1 \lambda(x)dx = 1$. Note also that $|s| \leq 1$ and it's easy to verify that $r_1(\cdot)$ and $r_2(\cdot)$ are uniformly bounded by 2. It follows that $C(r, s; \cdot)$ is uniformly bounded on $\Delta_{[0,1]}$.

Step 1: $K(f^A, f^B) = \sup_{\lambda\in\Delta_{[0,1]}} C(r, s; \lambda)$.

From Lemma 8, integrating by parts,

$$\begin{aligned} K(f^A, f^B)[m] &= m(1)s(1) - m(0)s(0) - \int_0^1 m'(x)s(x)dx + \int_0^1 [\alpha_m(x)r_1(x) + a_m(x)r_2(x)] \\ &= -\int_0^1 m'(x)s(x)dx + \int_0^1 [\alpha_m(x)r_1(x) + a_m(x)r_2(x)]. \end{aligned}$$

where we used $s(1) = 0$ and $m(0) = 0$. From above it follows that

$$K(f^A, f^B) = \sup_{\lambda \in \Delta_{[0,1]}} C(r, s; \lambda) \tag{19}$$

Step 2: $\sup_{\lambda \in \Delta([0,1])} C(r, s; \lambda) \geq \max_{u \in [0,1]} \{l(u), -s(u)\}$ ■

Consider the p.d.f. $\lambda_u \in \Delta_{[0,1]}$ defined by $\lambda_u(x) = 1/1 - u$ for $x \in [0, 1-u]$ and $\lambda(x) = 0$ for $x \in (1-u, 1]$. It is straightforward to verify that $C(r, s; \lambda_u) = l_1(u)$. Similarly, if $\lambda_u(x) = 1/1 - u$ for $x \in [u, 1]$ and $\lambda(x) = 0$ for $x \in [0, u)$, we have that $C(r, s; \lambda_u) = l_2(u)$. Finally, if δ_u is the Dirac delta distribution -an atom- on point u, we have that $C(r, s; \delta_u) = -s(u)$. It follows that $C^* \geq \max_{u \in [0,1]} \{l_1(u), l_2(u), -s(u)\}$.

Step 3: $\sup_{\lambda \in \Delta([0,1])} C(r, s; \lambda) \leq \max_{u \in [0,1]} \{l_1(u), l_2(u), -s(u)\}$.

We show this by solving an optimization problem closely related to (19). In particular we solve (19) with an additional constraint, namely, we restrict the feasible set to distributions that "simple", i.e., those that are piece-wise constant. We next use the fact that any distribution λ can be approximated by a sequence of simple distributions. More precisely, a probability distribution λ is said to be *simple* if there exists a finite grid $I^n = \{e_0, e_1, ..., e_{n-1}\}$ on $[0, 1]$ with $e_0 = 0 \leq e_1 \leq ... \leq e_{n-1} = 1$ such that $\lambda(x) = \sum_{j=0}^{n-1} \mu_j \chi_{[e_j, e_{j+1})}(x)$, where χ_E is the characteristic function for $E \subseteq [0, 1]$. ($\lambda(\cdot)$ is constant and equal to μ_j on each interval $[e_j, e_{j+1})$.) Note that since λ is a probability distribution, the numbers μ_j are non-negative and $\sum_{j=0}^{n-1} \mu_j (e_{j+1} - e_j) = 1$. The set of simple distributions with a fixed grid I^n is denoted by $\Delta(I^n)$. We characterize the value of the related problem

$$C^*(I^n) = \sup_{\lambda \in \Delta(I^n)} C(r, s; \lambda). \tag{20}$$

For a fixed grid, $\Delta(I^n)$ is a finite-dimensional space isomorphic to a simplex in Euclidean space. To see this, for $\beta \in \{1, 2\}$ define $r_{j\beta} \equiv \int_{e_j}^{e_{j+1}} r_\beta(x)dx$ and let $s_j \equiv \int_{e_j}^{e_{j+1}} s(x)dx$. For each vector $\mu = (\mu_0, ..., \mu_{n-1}) \in \mathbb{R}^n$ let

$$\Lambda(r, s; \mu) = \sum_{j=0}^{n-1} r_{j1}\alpha_j(\mu) + \sum_{j=0}^{n-1} r_{j2}a_j(\mu) - \sum_{j=0}^{n-1} s_j\mu_j \tag{21}$$

where $\alpha_j(\mu) = \min_{k\in\{0,..,j\}} \mu_k$ and $a_j(\mu) = \min_{k\in\{j,...,n-1\}} \mu_k$ are a "cumulative minimum" functions. Note that for a simple distribution $\lambda = \sum_{j=0}^{n-1} \mu_j \chi_{[e_j,e_{j+1})}$ we have that $C(r,s;\lambda) = \Lambda(r,s;\mu)$ and, thus,

$$\begin{aligned} C^*(I^n) &= \sup_{\mu\in\mathbb{R}^n} \Lambda(r,s;\mu) \qquad (22) \\ \text{subject to } \mu_k &\geq 0 \text{ and } \sum_{j=0}^{n-1} \mu_j(e_{j+1}-e_j) = 1. \end{aligned}$$

This multivariate optimization problem can be restated as a linear program. This is done by introducing the auxiliary variables $\alpha_j \equiv \alpha_j(\mu)$ and $a_j \equiv a_j(\mu)$, and the additional linear constraints $\alpha_{j+1} \leq \alpha_j$ (which derives from $\alpha_j(\mu) = \min_{k\in\{0,..,j\}} \mu_k$) and, similarly, and $a_j \leq \alpha_{j+1}$ (which derives from $a_j(\mu) = \min_{k\in\{j,...,n-1\}} \mu_k$). Note that the objective (21) is linear in the vectors of α's, a's, and μ's and so are the constraints. The solution of this linear program is attained by an extreme point of the polytope defined by the constraints. Using this it can be shown that (22) has μ^* of such that either (1) $\mu_k^* = \frac{1}{(e_{j+1}-e_j)}$ for some $k \in \{0,...,J\}$ and $\mu_j^* = 0$ for all $j \neq k$; (2) there exists $\widehat{j} \leq J$ such that $\mu_j^* = 1/J$ for $j \leq \widehat{j}$ and $\mu_j^* = 0$ for $j > \widehat{j}$; or else (3) there exists $\widehat{j} \leq J$ such that $\mu_j^* = 1/J$ for $j \geq \widehat{j}$ and $\mu_j^* = 0$ for $j < \widehat{j}$. We save the reader from details. Evaluating the objective at a point μ^* that satisfies condition (1) of the claim above we get

$$\Lambda(r,s;\mu^*) = -\frac{s_k}{e_{k+1}-e_k} = \frac{1}{e_{k+1}-e_k}\int_{e_k}^{e_{k+1}} (-s(x))dx \leq \max_{u\in[0,1]} -s(u).$$

Similarly, evaluating at point μ^* that satisfies condition (2) of the claim, it is straightforward to verify that $\Lambda(r,s;\mu^*) = l_1(e_{\widehat{j}+1}) \leq \max_{u\in[0,1]} l_1(u)$. Finally, for a point that satisfies (3) we conclude that $\Lambda(r,s;\mu^*) = l_2(e_{\widehat{j}+1}) \leq \max_{u\in[0,1]} l_2(u)$. We conclude that for any grid I^n, we have that $C^*(I^n) \leq \max_{u\in[0,1]}\{l_1(u), l_2(u), -s(u)\}$. Since this bound is independent of I^n, $C(r,s;\lambda) \leq \max_{u\in[0,1]}\{l_1(u), l_2(u), -s(u)\}$ for any simple distribution λ. To complete the argument it suffices to note that simple distributions are dense in $\Delta_{[0,1]}$: any distribution $\lambda \in \Delta[0,1]$ can approximated by a sequence $\lambda^n \to \lambda$ as $n \to \infty$ such that each element of the sequence λ^n is a simple distribution (See, for example, Theorem 2.4 in Stein and Shakarchi [2005]). Since $C(r,s;\lambda^n)$ is uniformly bounded (as argued above), by dominated convergence, $C(r,s;\lambda^n) \to C(r,s;\lambda)$ as $\lambda^n \to \lambda$. Since $C(r,s;\lambda^n)$ is bounded

by $\max_{u\in[0,1]}\{l_1(u), l_2(u), -s(u)\}$ for all n, we conclude that $C(r,s;\lambda) \leq \max_{u\in[0,1]}\{l_1(u), l_2(u), -s(u)\}$. This holds for any distribution λ, simple or not. The proof is complete.

A.5 Proof of Proposition 2

Consider the piece-wise linear measure M characterized by parameters λ and γ. Let $\Delta\pi^{AB} = \pi^A - \pi^B$, $\Delta L = L^A - L^B$ and $\Delta G = G^A - G^B$. Then $K(f^A, f^B)[M] = \frac{1}{T}\left(\Delta\pi^{AB} + \lambda\Delta L^{AB} - \gamma\Delta G^{AB}\right)$, which is linear in λ and γ. From section 2, the (LA) property holds as long as $\gamma + \lambda \geq 0$, and the monotonicity conditions (M1) and (M2) are $|\lambda| \leq 1$ and $|\gamma| \leq 1$. Condition (M3), which applies to $T \geq 3$, is $\phi'(d_+ - d) - \phi'(d - d_-) \leq 1$ for all $d, d^+, d^- \in [0,1]$ (ϕ differentiable everywhere except at 0). This translates into $|\gamma - \lambda| \leq 1$. Constraints (M1)-(M3) define a polytope and, it can be verified that the set of extreme points of this polytope is precisely E_T^l. Since $K(f^A, f^B)[M]$ is linear $K(f^A, f^B) = \max_{M\in M^l} K(f^A, f^B)[M]$ is attained by one of thr points in E_T^l.■

B Empirical Applications Appendix

B.1 Dynamic Attrition

If dynamic attrition in our sample is not random then our estimate $\widehat{f}$ of the poverty dynamics will be biased. We asses this concern following the procedure suggested by Fitzgerald, Gottschalk and Moffitt (1998). The scope of our analysis is limited to selection on observables because no proper instrument can be found for testing selection on unobservables.

Consider the parametric selection model.

$$P_t = P_{t-1}\alpha + \varepsilon_t \quad \text{if } A_t = 0 \tag{23}$$

$$A_t^* = Z_t\delta + \upsilon_t \tag{24}$$

$$A_t = \begin{cases} 1 & \text{if } A_t^* > 0 \\ 0 & otherwise. \end{cases} \tag{25}$$

In equation (23), $P_t \in \{0,1\}$ indicates whether the individual is considered poor at period t and ε_t is the error term. We are interested in estimating α, the proportion of poor population in $t-1$ that remains poor in period t. Now,

we only are able to run this regression for the subsample of non-attritors. Indeed, A_t takes the value of zero whenever the individual is observed in both periods and one otherwise. Equation (24) describes a selection model where A_t^* is the attrition latent variable, Z_t is an auxiliary variable observed for all units independently of whether they attrite or not (e.g. lagged or time-invariant variable) and υ_t stands for the error term.

Selection on observables is present if ε_t is independent of υ_t but ε_t is not independent of $Z_t|P_{t-1}$. Dependence of ε_t on $Z_t|P_{t-1}$ means that by including Z_t and P_{t-1} together in the first regression we should find a non-zero coefficient for Z_t. Independence of ε_t and υ_t implies that there is no unobserved variable affecting both the regression and attrition equation.

We develop two empirical tests for selection on observables. First we evaluate whether observed characteristics among attritors and non-attritors differ. We run a Linear Probability Model (LPM) with A_t as the dependent variable, taking value zero for non attritors and one for future attritors. In this case Z_t is a vector including gender, age, education, marriage status, race and relationship with the head of the household. Finding significant differences would suggest that attrition is not random, leading to a biased estimate of α in equation (23). [51]. The next table summarizes the estimates of the LPM.

Linear Probability Model for Future Attrition

[51] The LPM can also be used to obtain an unbiased estimation of α can be obtained by weighting the non attritor sample by the normalized inverse selection probabilitiesIn this case the proper weigh can be obtained from the linear probabiliy model using the following formula $w(Z_t, P_{t-1}) = \left[\frac{\Pr(A_t=0/Z_t,P_{t-1})}{\Pr(A_t=0/P_{t-1})}\right]^{-1}$. See, Fitzgerald, Gottschalk and Moffit (1998).

	U.S. 80s	U.S. 90s	Germany	Great Britain
Intercept	.126	.097	.171	.218
	(.005)	(.004)	(.007)	(.009)
Male	.047	.044	.039	.033
	(.005)	(.005)	(.006)	(.007)
Age	.002	.003	.004	.003
	(.0001)	(.0001)	(.0001)	(.0001)
Years of Education	.003	.004	-.002	-.004
	(.0006)	(.0006)	(.001)	(.002)
Head of Household	-.042	-.041	-.015	.007
	(.006)	(.006)	(.007)	(.009)
White	-.127	-.127	N/A	-.184
	(.005)	(.005)		(.000)
Married	-.028	-.052	-.034	-.058
	(.006)	(.005)	(.005)	(.009)
Sample Size	21137	25575	19917	14224
Number of Attritors	3103	3953	3361	2630
R^2	.036	.038	.041	.044

While each observable variable plays a statistically significant role explaining future attrition, the general finding is that observables do not explain much of the attrition process. R^2 is small in every case analyzed here, suggesting that people with similar characteristics experience different attrition behavior. This is consistent with what Fitzgerald et. al (1998) find for longer time frame using the PSID. The concrete implication is that estimated weights are close to one, which in turn suggests that selection on observables is not a significant source of bias.

The second exercise tests directly whether selective dynamic attrition biases the estimate of α. Here we require a third observation for each individual in order to compare the estimated parameter between the full and the non-attritors sample. We first run equation (23) for the full sample between periods t and $t-1$. Then we repeat the estimation on the restricted sample of those who do not attrite at $t+1$. If attrition present we should observe different coefficients. Note that this test also gives us the direction of the bias.

$$P_t = P_{t-1}\alpha + \varepsilon_t \quad \text{full sample} \tag{26}$$

$$P_t = P_{t-1}\alpha' + \varepsilon'_t \quad \text{if } A_{t+1} = 0 \tag{27}$$

If coefficients from equation (26) and (27) remain unchanged we conclude that the parameter of interest is not biased when the sample is restricted to those who do not attrite. The next table summarizes the estimation using the PSID sample. In column 1 of the next table we take $t+1$ as the period 91-95, t as 86-90 and $t-1$ as 81-85. In column 2 we define t as the period 91-95.

Testing for Selective Dynamic Attrition in the PSID

Dependent Variable= P_t	$t = 1986 - 1990$		$t = 1991 - 1995$	
Variable	Full Sample	Non-Attrite	Full Sample	Non-Attrite
Intercept	.051	.049	.059	.054
	(.002)	(.002)	(.002)	(.002)
P_{t-1}	.568	.562	.586	.583
	(.006)	(.006)	(.006)	(.006)
Sample Size	17081	15419	18302	15525
R^2	.375	.371	.358	.358

For $t = 1986 - 1990$ the estimated coefficient on P_{t-1} for the full sample is .568, while the same parameter for non-attritors is .562. The small difference suggests that poverty persistence is slightly lower for the non-attriting sample, however the difference is not statistically significantly different from zero. For $t = 91 - 95$. the estimated coefficients on P_{t-1} is .586 and .583 for the full and non-attrite sample. This evidence suggests that at for U.S. selective attrition does not introduce a significant bias.

The previous analysis requires that we observe at least three periods. In the case Great Britain we implement the same procedure considering the window 2001-2004 (in addition to the two windows for the 1990s decade). For Germany third window is the period 2002-2005.

Testing for Selective Dynamic Attrition in Great Britain and Germany

Dependent Variable= P_t	Great Britain		Germany	
Variable	Full Sample	Non-Attrite	Full Sample	Non-Attrite
Intercept	.025	.023	.096	.029
	(.002)	(.002)	(.002)	(.002)
P_{t-1}	.305	.297	.252	.279
	(.007)	(.008)	(.009)	(.007)
Sample Size	11140	9716	16122	12382
R^2	.13	.14	.11	.03

The previous table suggests that there is no selective attrition for Great Britain. In contrast, for Germany there is evidence of selective attrition. However, the data suggests that if anything, our estimation of poverty persistency for Germany on the non-attrite sample is downward bias. This would imply a lower ranking for Germany than the one inferred from the sample estimates. Hence, it doesn't affect the conclusion that Great Britain dominates Germany, and so does the U.S. if we use the headcount ratio as the underlying poverty status.

B.2 Statistical Inference with Inequality Restrictions

For the case of the Headcount Ratio proposition 1, $f^A \succeq f^B$ requires three inequalities to be jointly fulfilled. In our applications we observe the sample estimates $\hat{f}^A$ and $\hat{f}^B$ of the poverty dynamics f^A and f^B. The statistical inference associated to the joint inequality restrictions is done using a version of the Wald statistic suggested by Formby et al. (2004) and Kodde and Palm (1986). Let $\beta \in \mathbb{R}^3$ be the vector

$$\beta = \begin{pmatrix} f^A(0,0) - f^B(0,0) \\ f^A(0,0) + f^A(1,0) - f^B(0,0) - f^B(1,0) \\ f^A(0,0) + f^A(1,0) + f^A(0,1) - f^B(0,0) - f^B(1,0) - f^B(0,1)) \end{pmatrix}$$

and observe that the inequalities in Proposition 1 can be expressed as $\beta \geq \mathbf{0}$, where $\mathbf{0}$ is the zero vector. Testing the hypothesis $H_0 : f^A = f^B$versus $H_1 : f^A \succeq f^B$ is thus equivalent to $H_0 : \beta = 0$ versus $H_1 : \beta \geq \mathbf{0}$.

Let Ω denote the covariance matrix associated to β. Using the sample estimate $\hat{\beta}$ we compute a Wald statistic from the solution to the minimization problem

$$\min_{\gamma \in \mathbb{R}^3} (\hat{\beta} - \gamma)\Omega^{-1}(\hat{\beta} - \gamma). \tag{28}$$

Using $\bar{\gamma}$ to denote the solution, we can compute two Wald estimators

$$\begin{aligned} W_1 &= \hat{\beta}'\Omega^{-1}\hat{\beta} - (\hat{\beta}-\bar{\gamma})'\Omega^{-1}(\hat{\beta}-\bar{\gamma}) \\ W_2 &= (\hat{\beta}-\bar{\gamma})'\Omega^{-1}(\hat{\beta}-\bar{\gamma}). \end{aligned}$$

Let $W = \max\{W_1, W_2\}$ and $w = \min\{W_1, W_2\}$ and denote by cv and CV the lower and upper bounds of the critical value provided by Kodde and Palm (1986) for the relevant significant level and degrees of freedom. If $W > CV$ then H_0 is rejected. If $w < cv$ then H_0 is accepted. Finally if W_1 and W_2 fall between the cv and CV a Monte Carlo simulation is required to make an inference.

Solving (28) requires an estimate of covariance matrix Ω. Observe that $\beta = D\phi'$ where

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix}$$

and $\phi = (f^A(1,1), f^A(0,1), f^A(1,0), f^A(0,0), f^B(1,1), f^B(0,1), f^B(1,0), f^B(0,0))$. In particular, in this case the asymptotic covariances of the parameters are given by[52]

$$Cov(f(i,j), f(k,l)) = \theta = \begin{cases} \frac{\widehat{f}(i,j)(1-\widehat{f}(i,j))}{n} & \text{if} \quad i = k \ \text{ and } j = l \\ -\frac{\widehat{f}(i,j)\widehat{f}(k,l)}{n} & \text{else} \end{cases}$$

where n is the sample size. The covariance matrix of ϕ is

$$\Sigma = \begin{pmatrix} \theta^A & 0 \\ 0 & \theta^B \end{pmatrix}$$

from which the asymptotic covariance matrix of β is $\widehat{\Omega} = D\widehat{\Sigma}D'$.

For the case of the Poverty Gap we follow a similar procedure, but now the three inequalities required to rank $f^A \succeq f^B$ according to **??** are described by $\beta \in \mathbb{R}^3$

$$\beta = \begin{pmatrix} \pi_2^B - \pi_2^A \\ \pi_1^B - G^B - \pi_1^A + G^A \\ \pi_1^B + L^B - \pi_1^A - L^A \end{pmatrix}$$

[52] See Formby, Smith and Zheng (2004) for a detailed discussion about estimating covariance matrices for different type of transition matrices.

Again, testing the hypothesis $H_0 : f^A = f^B$ versus $H_1 : f^A \succeq f^B$ is equivalent to $H_0 : \beta = 0$ versus $H_1 : \beta \geq \mathbf{0}$. Solving (28) requires an estimate of covariance matrix Ω of β. Observe that $\beta = D\theta'$ where

$$D = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ -1 & 0 & 0 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

and $\theta = (\pi_1^A, \pi_2^A, G^A, L^A, \pi_1^B, \pi_2^B, G^B, L^B)$. We can estimate from the data the variance-covariance matrix $\lambda = Cov(\pi_1, \pi_2, G, L)$ for each society in order to estimate the covariance matrix of θ

$$\Gamma = \begin{pmatrix} \lambda^A & 0 \\ 0 & \lambda^B \end{pmatrix}$$

which yields the asymptotic covariance matrix of β, $\widehat{\Omega} = D\widehat{\Gamma}D'$.