The University of Manchester
Brooks World Poverty Institute

MANCHESTER 1824

Evaluating ten years of 'strategizing' for poverty reduction: A cross-sectional appraisal of the Poverty Reduction Strategy Paper (PRSP) initiative

Richard Marshall[1*]

Bernard Walters[2]

March 2011

BWPI Working Paper 143

Creating and sharing knowledge to help end poverty

[1] United Nations Development Programme (UNDP), Dhaka, Bangladesh

* Corresponding author

Richard.marshall@undp.org

[2] School of Social Sciences, The University of Manchester, UK

Bernard.Walters@manchester.ac.uk

Brooks World Poverty Institute
ISBN : 978-1-907247-42-2

Abstract

In late 1999 a joint meeting of the IMF and World Bank announced the introduction of Poverty Reduction Strategy Papers (PRSPs) as a means of securing comprehensive development, alongside a framework for the provision of increased financial support (specifically via debt relief). PRSPs were pathbreaking in at least two senses. First, because poverty reduction rather than other economic objectives became the focus of policy-based lending; and second, in bringing strategic planning back into the mainstream development agenda. As the decade progressed, PRSPs became the primary framework through which economic and social policy was crafted and managed in low-income countries, and there are now 67 PRSP arrangements in place. Yet, in spite of the passage of ten years, there remains no authoritative evaluation of the initiative's impact. This paper aims to fill this lacuna, by offering a cross-sectional appraisal based on sound counterfactual analysis. It makes use of a series of quantitative methods, including exhaustive econometric evaluations, of two specially constructed panel datasets. The analysis also employs Bourguignon's (2004) discussion of the poverty-growth-inequity triangle, and the role played by policy in mediating the core relations. The objective is to appraise performance in terms of poverty reduction, but also to disentangle the separate impacts of distributional change and economic growth. The results provide some evidence of a positive PRSP treatment effect in relation to poverty reduction, but with this operating exclusively via the growth channel. While this lends support to PRSPs as enhanced growth strategies, it undermines their claims to secure more widely balanced, and hence, pro-poor, growth. A number of evidential issues are also addressed, which cast doubt on the strength of the apparent performance gains.[1]

Keywords: Empirical studies of poverty dynamics, Evaluations and assessments of policies and programmes

Richard Marshall is Economic Adviser at United Nations Development Programme (UNDP) Bangladesh, Dhaka.

Bernard Walters is Senior Lecturer in the School of Social Sciences at The University of Manchester, UK.

[1] The views expressed in this publication are those of the author(s) and do not necessarily represent those of the United Nations, including UNDP, or their Member States.

1. Introduction

For the last decade, the Poverty Reduction Strategy approach, incorporated into Poverty Reduction Strategy Papers (PRSPs), has been the primary framework defining the relationship between the International Financial Institutions (IFIs) and low-income countries (LICs). Its objectives are, broadly, to prioritise and embed the objective of poverty reduction within LICs' economic development agenda, while simultaneously resolving, by inculcating a greater degree of national ownership of that agenda, the perennial principal-agent problem at the heart of the donor-recipient aid relationship. Many have hailed PRSPs as pathbreaking, as they identify poverty reduction as the primary goal of policy and re-habilitate the role of national planning in the development process (Booth et al., 2003).

The inception of the PRSP approach can be dated precisely to a joint World Bank/IMF meeting in 1999, which led quickly to its adoption and rapid deployment in 2000, so that 2010 marks its first decade. This provides an appropriate point for its evaluation, while ten years also means that a reasonable body of data on the performance of adopting and non-adopting countries has been accumulated to support statistical and econometric testing. The rhetoric of the process insists that PRSPs are individually structured to specific country circumstances, undermining attempts at identifying a general PRSP effect; this presumption has ensured that a relatively small number of evaluations have appeared in the literature to date. This paper rejects this view. Its aim is precisely to employ the emerging empirical evidence to provide an evaluation of the PRSP approach against the stated objective of poverty reduction. In particular, it employs a structured series of statistical and econometric studies, based on counterfactual comparisons between two specially constructed panel datasets of PRSP-adopting and non-adopting countries, to investigate whether any persuasive evidence at the aggregate level can be found, under PRSP arrangements, of enhanced poverty reduction, and its proximate drivers of economic growth and more equitable income distribution. Parenthetically, the paper also investigates the degree to which the data support the characterisation of the approach as simply the new face of old-style structural adjustment and stabilisation.

The genesis of the PRSP approach and these earlier studies are discussed in Section 2. This is followed in Section 3 by a brief description of the dataset to be used and the statistical and econometric approaches applied. Section 4 presents and discusses the results of a series of statistical cross-tabulations of differences in performance between PRSP and non-PRSP groups of countries, while Section 5 reports the results of two standard econometric approaches to panel data estimation – first differencing and fixed effects. Finally, Section 6 summarises the results and provides our conclusion.

2. The Poverty Reduction Strategy approach

In 1999, a joint meeting of the IMF's International Monetary and Finance Committee (IMFC) and the combined IFI Development Committee (DC) received, and agreed, two position papers on a development-planning framework that would rely on nationally authored Poverty Reduction Strategy Papers. The first of these reports, which was produced by the World Bank, proposed PRSPs as a new aid modality to govern the disbursement of its own concessional lending; the

second was a joint report on the next tranche of the Heavily Indebted Poor Countries (HIPC) debt relief scheme. This report proposed that PRSPs be the means for linking poverty reduction objectives to debt relief allocations; as such, PRSP-development and then adoption would serve as part of the pre-selection criteria. The two papers were closely coordinated, and outlined the basic shape of the arrangements, supportive framework and guiding principles, which have remained largely unchanged (see IEO, 2004; IEG, 2004).

Although the central objective was framed in terms of prioritising the poverty alleviation agenda, the approach was driven by the IFIs' needs to address the widely recognised failure of the conditionality driven approach to policy lending and to rebut increasing external criticism (see Christiansen and Holland [2003] and Marshall [2010] for a discussion of the detailed development of the PRSP approach). From 2000, PRSPs were rolled out rapidly, initially to HPIC qualifiers and progressively to all IDA members. Although the basic shape of these arrangements has changed little, there have been some modifications to the management procedures. Recent years have also seen some decline in the absolute primacy of PRSP programming within IFI operations. In 2009, the IMF renamed the Poverty Reduction and Growth Facility (PRGF) as the Extended Credit Facility (ECF) and broadened its scope. Meanwhile, although the World Bank has retained a very strong poverty focus, new priorities have emerged in the post-Wolfensohn years. Developments from the recipient side have been more significant. In recent years, a number of countries have either allowed their PRSPs to fall into abeyance or abandoned the process in favour of national planning frameworks; this suggests that the PRSP approach may have little long-term future as the primary framework governing the IFSs' relationships with LICS. Nevertheless, over the last decade, PRSPs were undoubtedly the primary means through which multilateral assistance was granted, and managed within LICs. The IFIs assert that that this has ensured national ownership of reform programmes, enabling the building of state capacity, while simultaneously securing a series of institutional improvements. These, along with additional gains – chiefly the refocusing of policy on poverty reduction rather than economic objectives and donor coordination gains – provide the case for a positive evaluation of their approach.

There have been a number of evaluation attempts of the PRSP approach, within which four distinct strands can be identified: the IFI's own periodic reports of the initiative's implementation; two extensive studies undertaken by the IMF's and World Bank's respective external evaluations bodies – the Independent Evaluations Office (IEO, 2004) and the Independent Evaluations Group (IEG, 2004); a set of independent studies which examine PRSPs' contribution to aid effectiveness; and a group of qualitative research studies based on multiple case studies.

However, for the purposes of this paper, most of these evaluations provide little guidance. As noted, the PRSP approach had a number of objectives, both from the perspective of the IFIs and also relative to the specific features of the PRSP process. The extent of the PRSP framework, and the complexity of issues covered by the stated intention to secure national ownership, mean that the notion of success is a very slippery one, with outcomes under different headings potentially judged against a wide range of possible objectives. These included, *inter alia*: institutional and governance reform; aid effectiveness; evidence-based policy development and implementation; broadening and extending national ownership of policy choices. Much of the evaluation literature addresses one or more elements from this set. However, the objective of this paper is to investigate whether the combination of changes inherent in the various PRSPs have, in the aggregate, contributed to a greater reduction in poverty than would otherwise be the case. This narrows the

range of relevant contributions considerably and, in fact, there are very few studies that even touch on this level of generality.

In addition, such an approach raises a question which has, hitherto, insulated the PRSP approach from quantitative evaluation: if the PRSP approach is individually tailored, then the possibility of identification of an aggregate effect due to PRSP-adoption is difficult to sustain and, in fact, the IFIs have argued for the non-testability of the programme in aggregate terms, because of the lack of an identifiable counterfactual. However, although it is difficult to identify a specific set of common policies across all PRSPs, this is not a fully convincing argument. First, there is sufficient commonality between different PRSPs to infer a strategic orientation with a corresponding policy template. Second, even if the Bank's description of PRSPs is accepted, if appropriate controls are employed it must be possible to test for the claimed benefits at the outcome level; if after ten years of application the PRSP does not appear to generate any observable change in the trajectory of poverty, then this surely raises some questions about its effectiveness, whatever other, possibly beneficial, impacts it may have. Central to the approach followed in this paper is the explicit assumption that PRSPs form an identifiable, and therefore testable, set of processes and/ or policy choices. As such, adoption of a PRSP is viewed as having a country-level treatment effect, which can be measured against a counterfactual benchmark given by non-adopting countries' performance.

In fact, despite the IFI's claim of non-testability, some elements of the reports complied by the IMF Independent Evaluations Office (IEO) and, to a lesser degree, the World Bank's Independent Evaluations Group (IEG)[2] offer some degree of aggregate evaluation in terms which are suggestive for the present study (IEO, 2004; IEG, 2004).

The IEO report does address the initiative's wider effectiveness and, in direct contradiction to the IFIs' assertions, tests outputs and initial outcomes via a set of counterfactual comparisons. Although limited, the comparative quantitative analyses provided remain the most substantial of any published source on PRSPs. The report presents two principal counterfactual analyses. The first is based on World Bank Country Policy and Institutional Assessment (CPIA) scores for International Development Association (IDA) adopters and non-adopters (repeated in the IMF's Operations Evaluation Department [OED] Review). The overall CPIA score improves less significantly for the PRSP group, while on the individual components, only the public sector management index records a better performance for the PRSP group. The second analysis employs IMF macroeconomic data for a group of PRGF and non-PRSP (but PRGF-eligible) countries to test for improved macroeconomic performance. These are also used specifically to examine claims of a disinflationary bias within PRSP arrangements. In this case, 23 PRSP adopters are compared against 26 non-PRSP but PRGF-eligible countries, for two three-year intervals (1997-1999 and 2000-2002). While both groups generally show overall deteriorations, the PRSP growth position weakens far less markedly. This performance appears all the more impressive in the light of the trade data, which show a substantial deterioration in the PRSP group's terms of trade. Also noteworthy is the greater improvement in the PRSP countries' external debt levels, though this is most likely a direct impact of reliefs granted under the HPIC initiative.

These data offer some *prima facie* evidence of better growth performance within the PRSP group, although the policy indicators tell a rather different story; there are only minor indications of better

[2] Formerly the Operations Evaluation Department (OED) of the World Bank.

fiscal management with revenues rising more significantly in the PRSP group, which are, however, offset by higher government expenditures, so that the overall deficit (within both groups) shows a similar level of deterioration. However, these figures are potentially consistent with greater ODA flows and better management (and, by implication, better allocation) of public spending within the PRSP countries. Indeed, the IEO use these data to support their finding that there is no evidence of a disinflationary bias or aid pessimism within PRSP arrangements. It is also worth noting that trade restrictiveness (measured here by the IMF's subjective index and the value of tariffs) declines considerably in adopting countries versus a static position in the non-PRSP group. In addition, the IEO report includes data from the case study materials (developed in concert with the IEG team), which show positive performance against national MDG targets, including poverty levels for a small sample of 12 adopting countries.

The IEG report (the World Bank's external evaluations body) adopts a less overtly critical tone and is informed, at least to some degree, by the Bank's non-testability defence. It explicitly notes the difficulties of counterfactual analyses and the limited amount of data available. The quantitative performance appraisal provided by the IEG report is also more limited than that offered by the IEO. It tends to avoid direct performance comparisons, holding to the Bank's line that policy objectives are nationally determined, and therefore, context specific. Nevertheless, the review does report on what it refers to as outcome performance, as recorded in annual progress reports (APRs) against national planning targets, and comparative material is also taken from the ten case study evaluations. The subsequent discussion is summarised under three headings: improvements in state capacity to deliver poverty reduction, the pro-poor orientation of policy responses and initial outcome performance.

In relation to the first, the review supplements the CPIA dataset reported above with case country experiences to show some improvements within PRSP countries over time. However, it also finds that the comparative performance (against non-adopters) has been poor. Like the IEO report, this shows that only on the public sector management category do PRSP-adopting countries perform better. The discussion also notes some improvement in overseas aid flows to PRSP countries between 2000 and 2003; aid levels in these countries are well above those enjoyed by non-PRSP IDA members. The increases were most substantial for the early adopters. The discussion is, however, limited and only graphical evidence is provided.

The second area, covering the poverty orientation of policy, is examined using the ten case study evaluations. The report's conclusions here are rather more positive, with the qualitative evidence suggesting that greater attention is being given to poverty objectives within strategic policymaking. Quantitative budgetary analysis is provided for 14 PRSP countries with reliable expenditure data, which, it is suggested, is supportive of better performance. The report finds that there has been a shift to what they refer to as poverty-reducing expenditures (PREs) – defined largely as allocations to the education, health, and social welfare sectors. Compared with the years prior to adoption, PREs in the PRSP group showed an average increase of 1.4 percent of GDP in 1999, and 3.9 percent of general government expenditure (GGE) between 1999 and 2003. However, it is evident that these increases are not statistically significant, and the upward trends were already well established in the pre-adoption period.

The analysis of poverty outcomes is still more circumspect. Here the report focuses on 12 PRSP countries that had provided annual progress reports (APRs) by 2004. Using the APR targets specified by individual national governments, and grouping these within the eight MDG categories, the IEG find that progress has been made for all goals, save for the child mortality maternal mortality targets. Yet the numbers of countries that responded, and the indicators used for each goal, vary considerably. Indeed, only three of the APRs directly reported (all positively) on the poverty reduction goal, and these three reports (out of a total of 14) are interpreted by the IEG as clear evidence of overall progress.

In summary, the present evaluation literature provides very few analyses that even address the question of the overall effectiveness of the PRSP approach. However, those few which – explicitly or implicitly – reject the insulating claim of non-testability, provide some evidence of improved growth performance, reject the claim of disinflationary bias and, more circumspectly, provide some evidence of improved institutional, expenditure and other changes likely to impact positively on poverty outcomes. However, it is clear that the evidence is patchy and inconclusive, justifying the attempt to construct a formal evaluation based on newly available data.

3. Research approach and data

The initial investigation of the data is in term of standard statistical tests between a treatment and control group of countries. This is followed by more thoroughgoing econometric examinations, which enable heterogeneity and selection biases to be addressed. It is, however, important to be clear about the specific objectives, and the limitations imposed by the data environment in which the analysis is carried out. The central question being examined is whether better poverty outcomes are associated with, and can be traced to, PRSP adoption. In addition, the analysis attempts to investigate the subsidiary question of whether the PRSP framework represents simply a vehicle for implementing traditional stabilisation policies.

The overall framework used to test for the impact of PRSP adoption is the growth-inequality identity that poverty reduction (the end goal of PRSPs) is the product partly of growth in average incomes, and partly of the change in the distribution of income. This lies at the heart of the poverty decomposition methodologies developed by Datt and Ravallion (1992) and by Kakwani (1997) and their application to poverty dynamics by Bourguignon (2002 and 2004).

Formally, this decomposition result, described by Bourguignon (2002) as '*the standard relation*', can be specified in terms of a change in any additive poverty measure and a fixed poverty line. This is expressed below, for the change in a poverty measure (P) over two time periods ($t, t\text{-}1$); F depicts the income distribution function; and z a real terms constant poverty threshold.

$$\Delta P = P_t - P_{t-1} = F_t(z) - F_{t-1}(z) \qquad [1]$$

If we normalise by the average income ($\bar{y}$), the change can be written as the decomposition of the two effects, in which the terms are independent of the scale of incomes:

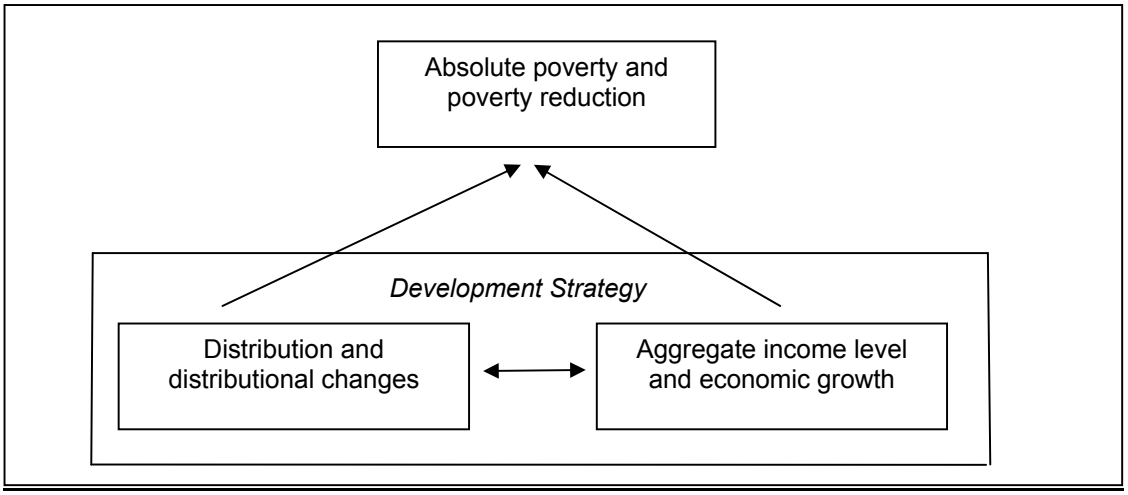
$$= \left[F_t\left(\frac{z}{\bar{y}_t}\right) - F_t\left(\frac{z}{\bar{y}_{t-1}}\right)\right] + \left[F_t\left(\frac{z}{\bar{y}_t}\right) - F_{t-1}\left(\frac{z}{\bar{y}_t}\right)\right] \qquad [2]$$

Thus, in equation 2, the first square bracket represents the income or growth impact – the change holding distribution constant (ΔP_G), the second the distributional effect where income is held constant (ΔP_I).

$$\Delta P = \Delta P_G + \Delta P_I \qquad [3]$$

Bourguignon (2004) argues that aggregate changes in the poverty level, which are a function of growth in incomes, the initial distribution of income and distributional change, can be influenced by policy choices, institutions and structures. Moreover, he describes this process in terms of an interlocking relationship – the poverty-growth-inequality triangle (see Figure 1). Poverty reduction is, therefore, maximised through enhanced output alongside the effective management of inequality. Thus policy, and in the context of this paper, PRSP adoption, is a conditioning variable on the standard relation, and in the analyses that follow, adoption is modelled as a treatment effect. Evaluation is based on counterfactual comparisons between treatment (PRSP) and control (non-PRSP) groups of countries.

Figure 1: Bourguignon's (2004) poverty-inequality-growth triangle



The analyses which follow make use of two panel datasets (one based on national poverty lines and the other international lines) comprising 68 PRSP-adopting and non-adopting groups of countries over a 12-year time frame running from 1996. Appraisal techniques tested for differences in outcomes between the two groups; first, by way of statistical tests of difference of means (for outcome variables); and second, via econometric evaluations, in which the treatment effect was identified by the coefficient on a binary variable. Within each of the evaluations, poverty is represented by the headcount measure, distribution by the Gini coefficient, and incomes (and hence growth) by an index of real terms US dollars per capita.

In recognition of the heterogeneity of the data, the statistical testing makes some use of matching techniques, partially using Propensity Score Matching (PSM) methods. Changes are also annualised to allow for the large variations in the length of the poverty reduction episodes. The econometric approaches address the challenges posed by unobserved heterogeneity directly within the specifications employed, but vary somewhat in the model used and the form of the two datasets

employed. The First Differences (FD) estimator models poverty change in terms of contiguous changes. In order to achieve sufficient balance in the panel, the 12 years of data were parsed into four blocks and averaged. In contrast, the Fixed Effects (FE) estimator models the relation in levels as a pooled cross-section for the full 12-year sample. Instrumental variables (IV) techniques are also employed to control for endogeneities arising from the non-random nature of PRSP adoption.

A major issue to be resolved is the operational definition of poverty. In the interest of clarity and data availability, the headcount ratio was selected. However, two alternative headcounts presented themselves: those given within national surveys undertaken primarily by government statistical agencies; and international dollar-a-day poverty dataset held by the World Bank.[3] On first inspection, the former would appear to fail a key test of comparability across countries, and potentially that of consistency through time. In line with other researchers (see for example, Ravallion and Chen, 1997; and Chen and Ravallion, 2008), international lines would appear to be the preferred choice.

However, such a conclusion overlooks a number of important issues. Foremost, a separate critical literature has developed over the quality and consistency of international poverty lines, which questions their claimed superior longitudinal and cross-country comparability. Criticisms centre on disquiet over the Purchasing Power Parity (PPP) adjustment, and that any potential measurement errors within the Living Standards Measurement Survey (LSMS) modality would also affect the comparability of the dollar-a-day series. There is also a definitional problem, in that the dollar–a-day line, inevitably, is somewhat arbitrary and has a weak correspondence with welfare (Reddy and Pogge, 2005). A basic causal issue also needs to be noted. Given that PRSPs target nationally generated poverty data, it would be rather perverse to track their performance using an alternative metric. In recognition of these challenges, and the benefits accruing from the triangulation of results, two panels (national and international) are employed in the analyses reported below.[4]

The primary means of ensuring the comparability within the national panel was a strict filtering exercise to ensure that all data records were of a sufficiently high standard (benchmarked to the LSMS series) and rooted in a common calorific (2,100 calories a day) threshold.[5] This required building the panel from primary sources (LSMS publications, World Bank Poverty Assessment and national poverty reports) for the 12-year period from 1996 to 2007. The dollar-a-day line data were sourced from the World Bank's Povcalnet database[6] (an augmented form of the data held by the Word Development Indicators series).

Both datasets include the variables required by the standard model (poverty levels, per capita incomes and Gini coefficients). Additionally, PRSP status (year and length of adoption), and the basis of the poverty estimates (income versus consumption) were recorded within each. Further variables were incorporated within the three separate formats of each dataset, which were shaped to meet the requirements of the different stages of the analysis. Additional variables included IFI indebtedness (employing various lags) to allow for the use of IV methods, and structural adjustment

[3] It is worth noting that these two sources often rely on the same survey data – via national poverty surveys supported by the World Bank-funded Living Standards Measurement Survey (LSMS) programme.

[4] Full details of the data and the process of compilation are available, on request, from the authors.

[5] Given in terms of meeting the practices adopted within the World Bank's Living Standards Measurement Survey (LSMS) series.

[6] http://iresearch.worldbank.org/ROOT/body.htm

proxies within the statistical testing data to investigate the presence of deflationary biases within PRSP arrangements. The panels included data for a 12-year period from 1997, and some cleaning of the datasets took place, including some limited use of interpolation for missing values.

Data availability was a key issue during the compilation process. The initial objective was to include all 63 PRSP adopting countries within the panel; yet, astonishingly, only 29 had undertaken two or more high quality poverty surveys. This in itself is a major finding, underlining the widespread inadequacy of monitoring and evaluation resources in PRSP countries. The data from adopting-countries were supplemented (in each dataset) with 34 non-adopting countries to form a control group. Selection of this group was also conditional on data availability, again introducing the possibility of non-comparability, but efforts were made to balance the sample, both geographically and in terms of economic characteristics. Moreover, the inclusion of data prior to the introduction of the initiative (in 2000), permitted some same-country records to be employed, thus considerably boosting the number of records. Given some of these included PRSP countries prior to adoption, this effectively means the basis of comparison is a hybrid with-without and before-after counterfactual approach, as adopters' performance is compared with both non-adopters and their own data prior to adoption.

The varying requirements of the different stages of analysis necessitated the compilation of three formats of the two panels, and thus, six datasets in all were compiled. The key issue was the need to allow for the sporadic nature of poverty measurement, which, at best, followed the three-to-four-year LSMS cycle, and therefore the presence of a large number of missing records. Allied to this, efforts sought to allow for the representation of time effects and to get the maximum out of the data available. The three formats of the panel data were as follows:

- For the statistical testing, each country's record was divided into poverty reduction episodes (as in Bourguignon, 2002). Episodes ranged between two and seven years, the annualised change was then calculated for all of the variables of interest. This process yielded 83 observations for the national and 143 for the dollar-a-day panel.
- For the First Differences (FD) estimator, the datasets were *parsed* into four three-year periods. Annual observations within each period were averaged and some interpolation was necessary to provide a contiguous balanced panel. This gave 75 records for the national panel and 129 for the dollar-a-day panel.
- For the Fixed Effects (FE) estimator, the core 12-year dataset was retained with very minor modifications. This yielded the largest number of observations: 175 for the national and 256 for the dollar-a-day panel.

An additional set of tests was included within the initial statistical evaluation to examine the impact of PRSP adoption and stabilisation variables. At each stage of analysis, standard restricted samples were also modelled: at the testing stage by imposing a common support restriction (derived from an estimated propensity score, discussed below); and at the regression stages by excluding income-based survey data. Throughout the analysis, a lead-in time of two years was allowed for the time between PRSP adoption and the treatment impact to emerge. This period was chosen as the maximum lag, given the limited data availability.

4. Statistical evaluation

This section presents and discusses the results of standard cross-tabulations (and the associated statistical tests) for changes in the principal variables (poverty, growth and inequality) against PRSP status. In essence, these comparisons provide a difference in differences evaluation of outcomes between adopting and non-adopting groups of countries. This initial stage of the evaluation seeks to highlight *prima facie* evidence of performance differences and to identify any relationships requiring further investigation. In addition, the tests also examine the channels through which any PRSP poverty reduction benefit may be operating (via growth in incomes and/or distributional changes) and the argument that PRSPs merely represent a re-packaged form of structural adjustment.

The tests are in the form of comparisons of the mean annualised change in poverty, per capita incomes and inequality of the treated (PRSP) and control (non-PRSP) groups of countries for poverty reduction episodes (a period of two to seven years bounded by two comparable poverty surveys) within the 12-year sample interval. Any statistically significant differential serves as an approximate estimate of any treatment effects (ATE).[7] In order to improve comparability, tests using a restricted control group (derived using a propensity score method) were also implemented for each variable.

The tests rely on the episode-based format of the two panels: with 83 poverty reduction episodes in the national dataset (26 treatments against 57 control observations); and 143 episodes in the dollar-a-day dataset (30 treatment and 113 control observations). The key statistics provided within the tables are the mean changes and accompanying t-statistics. These are given for the one-sided test (i.e. the treatment mean is greater or less than that of the control group), and its double-sided counterpart (i.e. that the difference between treatment and control mean outcomes is not zero). The former is the primary basis for establishing a significant relationship, as the purpose is to test for improved performance within the treatment group. Formally, the null hypothesis is defined as one of no difference or worse PRSP performance and the alternative as better PRSP group performance. For example, for poverty, if *d* is defined as the difference between the change in the poverty levels between the control, non-PRSP group and the PRSP adopters, then the test is H_0: $d \leq 0$ against H_a: $d > 0$.

The tests for supplementary variables were framed in a similar manner, with better PRSP performance captured by rejection of the null hypothesis. Test results are reported for both the full and a restricted sample. The use of a restricted and better-matched sample has the potential of reducing the impact of selection biases and the level of heterogeneity, with clear benefits for both the accuracy and power of the tests. Propensity Score Matching (PSM) methods were used to identify those control group countries with similar characteristics to PRSP adopters, and thus to restrict the test data. The control group within the restricted test is given by those counties falling within the area of common support (the range in which the two propensity score distributions overlap).

Table 1 provides a summary of the test results for poverty reduction outcomes (ΔP), economic growth (represented by the change in per capita income, ΔPCY) and distribution (captured by

[7] The ATE here is given as the difference between the group averages for treatment (PRS) group and the control (non-PRS) group performance.

Table: 1: Cross-tabulation of poverty reduction and PRSP status

1	2	3	4	5	6	6
	Category	Non-PRSP mean change [standard deviation]	PRSP mean change [standard deviation]	Overall mean change [standard deviation]	Difference in means [t statistic]	Probability Diff>0 [Diff≠0]
(a) ΔP	National panel - Full sample - Restricted	 -0.90 [2.99] -1.09 [2.93]	 -1.75 [3.22] -1.75 [3.22]	 -1.17 [3.07] -1.39 [3.06]	 0.85 [1.14] 0.66 [0.79]	 0.13 [0.26] 0.22 [0.43]
	Dollar a day panel - Full sample** - Restricted*	 -0.34 [2.19] -0.37 [3.20]	 -1.49 [2.70] -1.49 [2.70]	 -0.58 [2.33] -0.80 [3.04]	 1.16 [2.17] 1.12 [1.67]	 0.02 [0.04] 0.05 0.10]
(b) ΔPCY	National panel - Full sample*** - Restricted**	 2.45 [3.70] 3.02 [4.09]	 5.38 [5.24] 5.38 [5.24]	 3.36 [4.43] 4.11 [4.77]	 -2.94 [-2.93] -2.36 [-1.86]	 0.00 [0.00] 0.03 [0.07]
	Dollar a day panel - Full sample* - Restricted	 3.29 [4.49] 3.38 [5.41]	 4.88 [5.05] 4.88 [5.05]	 3.63 [4.64] 3.95 [5.30]	 -1.59 [-1.57] -1.50 [-1.25]	 0.06 [0.12] 0.11 [0.22]
(c) ΔG	National panel - Full sample - Restricted	 0.05 [1.25] 0.11 [1.40]	 0.06 [1.41] 0.06 [1.41]	 0.05 [1.29] 0.09 [1.39]	 -0.01 [-0.04] 0.05 [0.14]	 0.48 [0.97] 0.55 [0.89]
	Dollar a day panel - Full sample - Restricted	 0.03 [1.19] 0.07 [1.56]	 0.12 [1.26] 0.12 [1.26]	 0.50 [1.20] 0.09 [1.44]	 -0.09 [0.37] -0.50 [-0.15]	 0.36 [0.72] 0.44 [0.88]

Note: Significant results are indicated in the tables by single, double or triple asterisks for the ten, five and one percent levels, respectively.[8]

Source: Authors' calculations.

[8] Prior to testing, variance comparison tests were undertaken to determine whether an unequal variance procedure was required (Satterthwaite's approach was used here).

changes in the Gini coefficient, ΔG), for both the full and restricted samples and for both national and dollar-a-day panels. The data for the non-PRSP (i.e. the control) group is given in column 2 and that for the PRSP (i.e. treated) group in column 3. The differences in outcomes and the test statistics are given in columns 6 and 7, respectively.

The results for the change in poverty suggest a positive impact of PRSP adoption, with a clear (positive) differential in the dollar-a-day panel in both the full and restricted samples (significant at the five and ten percent levels, respectively). However, the difference is not significant in the national dataset (with or without restriction). The estimated ATE in the dollar-a-day sample amounts to a further annual reduction of around 1.1 population percentage points in the headcount (in both samples – see block a column 6). Clearly, this represents a substantial effect, although it is important to note that the standard deviation between individual country records remains high, at around 2.0 to 3.0 population percentage points per annum.

Test data for differences in growth performance provide more persuasive evidence of post-adoption gains (see block b). The superior performance in the treatment group is sizeable and statistically significant in both national and dollar-a-day panels. In this case, the differential and the significance level are stronger in the national dataset. Within the full sample, the annualised mean gain in the index is 2.9 points in the national panel versus 1.6 points in the dollar-a-day panel (column 6). Similarly, the level of significance within the national panel reaches the one percent level in the full, and the five percent level in the restricted sample, against only the ten percent level in the dollar-a day panel. The level of variation, although high, is considerably less than that for the poverty data. Clearly, better growth outcomes underpin the poverty reduction reported above. However, the apparent lack of any feedthrough to the national data requires further consideration. Two alternatives present themselves: either growth is having no impact on national poverty rates; or national poverty thresholds are not consistent through time. Ravallion (1994), in arguing for the use of international poverty lines for cross-country comparisons, notes the tendency for real terms growth in the value of national poverty lines. It is possible that this is reflected in these results.[9]

Examination of the test results for the impact of adoption on distributional change reveals no discernable difference in performance between adopting and non-adopting country groups (see block c). Moreover, although insignificant, the statistics for the dollar-a-day panel show deterioration in Gini coefficients in the PRSP group relative to the control group. Additionally, the level of variation is exceptionally high in both datasets.

As a whole, these results underline the importance of growth as the channel through which gains in poverty reduction appear to have been secured within PRSP countries. There is little evidence that PRSP adoption has improved the management of inequality within the development process, or that growth has been any more pro-poor under PRSP arrangements (in a *Strong* sense).[10]

An identical testing procedure was used to examine the claim that PRSPs merely represent a facade for IFI-sponsored structural adjustment policies. The tests here made use of a similar, but separate, 12-year dataset for the same 68 countries, which included balance of payments and inflation data in place of poverty, growth and inequality variables. Several tests were carried out to

[9] This should, however, be somewhat mitigated by the use of LSMS data.

[10] A '*Weak*' position regards any poverty-reducing growth as pro-poor; in contrast the '*Strong*' position holds that the poor must benefit more than proportionately.

examine the impact of PRSP adoption on stabilisation variables (taken as hallmarks of the initial impact of structural adjustment policies). These included comparisons of the current account, the trade balance, Consumer Prices Index (CPI) and GDP deflator, specified in levels, in differences

and in variations against the mean. The results for the preferred test, that of divergences from national means for the current account balance and the Consumer Prices Index (CPI) are reported in Table 2.[11] The hypotheses being tested are that PRP adoption is associated with improved stabilisation outcomes (i.e. negative variations against the mean in both cases).

Table 2: Cross-tabulation of stabilisation variables against PRSP status

1	2	3	4	5	6
Category	Non-PRSP deviation from period mean [Standard deviation]	PRSP deviation from period mean [Standard deviation]	Overall deviation from period mean [Standard deviation]	Difference in means [t Statistic]	Probability Diff>0 [Diff≠0]
Current account balance as % of GDP***	0.33 [4.13]	-1.29 [4.92]	-0.00 [4.35]	1.62 [4.23]	0.00 [0.00]
Consumer Prices Index (CPI)	-0.64 [43.3]	+2.30 [5.45]	-0.00 [38.45]	-2.94 [-0.89]	0.81 [0.37]

Source: Authors' calculations.

The results of these tests are striking, and strongly reject the claims made. In the case of the balance of payments, the results show a highly significant variation (at the one percent level) between the two groups. However, contrary to expectation, and the critical arguments, the PRSP group performs worse than the non-PRSP group – the adopting group's average current account deficit worsens (relative to the mean) by around 1.3 percent of GDP (column 3), whereas the latter records a mild improvement (column 2). The test data for the CPI are not as decisive, with no significance attached to the results. Yet these also suggest that, at the aggregate level, PRSP arrangements are free from the kind of overt policy biases claimed by some authors.

Finally, it is important to reflect on the adequacy of the testing approach and the sensitivity of the results. Three major criticisms can be made: fragility of the test statistics, due to the low sample sizes and the quality of the data; unobserved heterogeneity within the data, as underlined by the high level of variation within the entire test samples; potential endogeneity within the data

[11] These were preferred, as they show the variation against longer-term trends under control and treatment groups for the most commonly accepted measures of internal and external balance.

generation process at work. PRSP adoption is a non-random event and arises out of a series of pressures operating on low- and middle-income countries. These include dependency on the IFIs and the effective conditionality of concessional credit lines, notably HIPC debt relief. There is a very basic difference between adopters and non-adopters, and thus the counterfactual employed does not approximate a natural experiment. These basic statistical methods offer no means of dealing with these biases.

However, in spite of these challenges, the testing exercise is helpful in highlighting the main issues to be addressed. Four key points can be made. First, the test results offer some *prima facie* evidence that superior poverty reduction and growth performance is associated with PRSP adoption, albeit, this is confined to the dollar-a-day panel. Second, it is clear that declining poverty within the PRSP group is supported by better growth outcomes, and that there no discernable distributional improvement. Thus, third, there is little evidence that growth is any more pro-poor under PRSP arrangements and this contradicts a core policy rationale of the Initiative. Fourth, and more positively, there appears to be no substance in critical claims that PRSPs have a policy bias towards stabilisation outcomes. In sum, although the testing evidence is far from conclusive, there is some evidence of a PRSP effect, but this also requires further and more thoroughgoing investigation.

5. Regression-based evaluations

This section presents and discusses the results of regression-based evaluations of PRSP performance. The objectives are to address the limitations of the statistical tests reported in Section 4, but also to extract additional information from the data. Two regression approaches, First Differences (FD) and Fixed Effects (FE), are employed to identify and evaluate any treatment effects. The model specified, under both approaches, is based on the standard relation linking poverty reduction to growth in incomes and distributional change. Informed by Bourguignon (2004), policy choices are taken to influence and mediate this relation, and thus both approaches estimate the impact of PRSP adoption on poverty (in differences and levels) as a binary treatment variable. The relation is examined using both national and dollar-a-day datasets. The binary adoption variable is identified only where a national PRSP has been in place for at least two years.

The two regression approaches necessitated some reshaping of the panel datasets:

- First Differences (FD) estimation relied on parsed versions of the original datasets into four equal three-year periods, covering the 12-year period from 1996. This yielded two contiguous, contemporaneous, and therefore fully balanced, panels of 76 countries in the national dataset and 129 countries in the dollar-a-day dataset.
- Fixed Effects (FE) estimation simply used the original annual datasets for the full 12-year period as two pooled cross-sections. This gives two unbalanced panels of: 175 records and 58 groups in the national dataset; and 256 records and 63 groups in the dollar-a-day dataset.

These standard panel methods have the advantage of removing non-varying unobserved effects from the basic relation being tested. In the FD model, this is achieved through differencing, and in the FE model through a time de-meaning process.[12]

These specifications were modelled using the standard relation: poverty (given by the Headcount ratio) is the dependent variable *(P0)* and per capita incomes (*PCY*) and distribution, given by the Gini Coefficient (*G*), are the independent variables (specified in differences or levels); with binary dummies added for adoption (PRSP) and time effects (*Pr3 and Pr4* in the FD regressions and *Yr2 to Yr12* in the FE regressions). The FD estimator is the preferred specification, as it more naturally captures the policy-performance hypothesis being examined, and the dynamic form of the standard relation. However, given the unbalanced nature of the original panel datasets, the FE estimator is more efficient in its use of the available data. To an extent, therefore, the discussion employs the FE results as a means of validating the former.

The main regression results (as with the statistical tests) offer a hybrid before-after and with-without counterfactual comparison, where PRSP-adopters are compared against observations for non-adopting countries, and their own data prior to adoption. A significant (negative) coefficient on the PRSP dummy is taken as evidence of a beneficial PRSP treatment effect on poverty outcomes. Where these were found, further efforts were made to identify whether these resulted from higher growth and/or a more equitable distribution of incomes. This was achieved by interacting the PRSP dummy with the respective per capita income and Gini coefficient variables and re-estimating the regressions.[13]

To address potential endogeneity biases, Instrumental Variables (IV) techniques were applied to the base specifications. These biases have a number of sources, but are likely to arise, mainly from the non-random nature of selection into treatment, and hence, the presence of simultaneities. The level of indebtedness to the concessional financing facilities of the IFIs, in the current period, and lagged to reflect the position prior to adoption, were used as instruments for adoption. The intuition behind this is the parallel requirements of the HIPC scheme, for which adoption is a decision point condition, and thus participation is separately conditioned by pressures to secure debt relief. Two instruments were employed – the current level of all concessional IFI finance outstanding as a share of GDP, reflecting the post-adoption impact of debt relief, and second, the level prior to adoption, to reflect the position some three to four years previously (made up of the two year lead-in time and the PRSP preparation period of one to two years). The expectation is that, while the latter would reflect a strongly positive correlation with PRSP status, the former would be negative in differences and only weakly positive in levels.

A standard two-stage procedure was used to estimate the IV results, with the fitted values of the first stage (the probability of adoption) being substituted for the binary PRSP adoption variable in the second regression. However, prior to this, regressions were carried out to test the strength of the correlation between the debt variables and adoption, and thus their validity as instruments. In

[12] See any standard intermediate econometrics text, for example, Wooldridge (2003).

[13] Efforts were also made to test for differences between early and later PRS-adopters and HIPC and non-HPIC status. However, the limited sample size made this line of inquiry impractical.

the event, this requirement was only satisfied in the case of the FD models. These were augmented by auxiliary regressions, and standard test statistics, to determine the presence of endogeneity.

The analysis yielded two sets of results (FD and FE) for the two (national and dollar-a-day) panels. In the case of the FD results, these were further divided two ways, the base OLS and IV regressions. The base estimates were also presented with restricted sample results comprising only consumption-based data. This is in recognition of the fact that the full panel includes a number of income-based surveys, so altering the parameters within the standard relation. The FD IV results were compared against a comparable base OLS regression. The models were also subject to a standard suite of tests.[14] Given the presence of non-normality in the FD estimators and some evidence of heteroscedasticity in both approaches, robust standard errors were used throughout.

The basic FD model is based on a standard relation expressed in differences (∆) form (for poverty, income and distribution). In order to allow for minor variations around the average three-year period, changes were also annualised.

$$\Delta P0 = \alpha + \beta_1 \Delta PCY + \beta_2 \Delta G + \beta_3 PRS + \beta_4 Pr3 + \beta_5 Pr4 + u \quad [4]$$

The two-stage IV estimation procedure relies on IFI indebtedness, both current (IFID) and lagged by one three-year period as instruments for PRSP adoption ($IFIDp_{-1}$) (this approximates the likely four-year interval referred to above). The first stage is estimated as a linear probability (LPM) model specified in differences:

$$First\ stage: \mathbf{Prob}\ (PRS) = \alpha + \beta_1\ \Delta IFID_p + \beta_2\ \Delta IFID_{(p-1)} + \beta_3 \Delta PCY + \beta_4 \Delta G + \beta_5 PRS + \beta_6\ \mathbf{Pr4} + u \quad [5]$$

$$Second\ stage: \Delta P0 = \alpha + \beta_1 \Delta PCY + \beta_2 \Delta G + \beta_3 \widehat{PRS} + \beta_4 \mathbf{Pr4} + u \quad [6]$$

It is important to emphasise that the datasets for the IV regressions (and the counterfactual basis) are different from the main regressions. Here the data are restricted to post-2002 observations – this is because PRSPs did not exist prior to 2000 and there is the two-year lead time used in identifying post-adoption effects. The counterfactual being tested is, therefore, purely a with-without comparison. In order to maintain comparability, the base OLS regression is re-estimated on the same basis. Hence also, in the model, only one time period dummy is included. The results for the base OLS regressions are provided in Table 3 and the IV results with the comparable OLS output are given in Table 4. The overall picture to emerge is significantly less supportive of a PRSP impact than the results reported in the previous section.

In the base OLS regressions: the national panel results merely find support for the standard model (Table 3, columns 2a and 2b). The per capita income variable is significant and strongly so (at one percent level) in the full and restricted (consumption only) regressions, both with a coefficient of around 0.2. The Gini coefficient is also close to being significant in the full base regression (column 2a). The treatment (PRSP) dummy is not significant in either regression. The R^2 value is around 0.20 and is significant at the five percent level in the base, but dips somewhat in the restricted, regression. Time effects are not found to be significant.

By contrast, the dollar-a-day results find PRSP adoption to be strongly significant at the one per-cent level in both the base and restricted OLS results (columns 3a and 3b). The adoption dummy

[14] Full regression output is available from the authors on request.

records a negative coefficient of 1.9 and 2.3, respectively. This implies that the marginal effect of adoption, on average, drives down the headcount ratio by around two population percentage points per annum. The overall explanatory power of the model (with an R^2 value of 0.12) is lower than in the national dataset, but of similar significance. Again, no evidence of any time effects is present. A further set of iterations of the base regressions made to test for the impact of varying the PRSP lead-in time, found these results (including the PRSP performance gain) to be fairly stable.[15]

Table 3: Summary results for First Differences (FD) base OLS regressions

1	2a	2b	3a	3b
FD of variable [t & F statistic] Dependent variable = poverty rate	National data		Dollar-a-day data	
	FD OLS	FD OLS Ex. income	FD OLS	FD OLS Ex. income
Per capita income	-0.2339*** [-3.65]	-0.1860*** [-3.25]	-0.0817 [-1.23]	-0.0976 [-1.04]
Gini coefficient	+0.4576 [+1.59]	+0.3986 [+1.32]	+0.3393 [+1.36]	+0.2974 [+1.02]
PRSP adoption	-0.6018 [-0.85]	-0.1198 [-0.17]	-1.8752*** [-2.79]	-2.360*** [-3.10]
Period 3 dummy	+0.2797 [0.39]	-0.2583 [-0.36]	+0.5656 [+0.98]	+0.9459 [+1.11]
Period 4 dummy	+1.4362 [1.41]	+0.7932 [+0.86]	-0.0574 [-0.10]	+0.3823 [+0.50]
R squared	0.2105** [3.16]	0.1617** [2.57]	0.1200** [2.43]	0.1262* [2.07]
Observations	76	66	129	83

Source: Authors' calculations.

The IV results given in Table 4, in contrast, reveal a rather different picture. After instrumenting for PRSP adoption, there is no evidence of a performance benefit in either panel. As shown in the Table, the IV and comparable OLS regressions for the national panel simply support the standard relation (columns 2a and 2b). More importantly, within the dollar-a-day panel, the PRSP impact, which is evident (at the one percent level) in the comparable base regression with a coefficient of 1.97 (column 3a), is not present in the IV results (column 3b). These findings represent a major qualification, and cast doubt on the treatment effect identified in the base FD regressions. The R^2

[15] Further analysis shows the PRS effect remained significant, with a lead-in time of one year, but was not apparent with no lead-in period.

values for the national IV regressions are significant (at the one percent level). Although the IV regression for the dollar-a-day data is statistically weak, the first stage regression shows the instruments are both significant and valid. Moreover, follow-up tests suggest that endogeniety is present in the dollar-a-day model (though not in the national dataset).[16]

Table 4: Summary results for First Differences (FD) IV regressions

1	2a	2b	3a	3b
FD of variable [t & F statistics] Dependent variable = poverty rate	National data		Dollar-a-day data	
	Comparable FD OLS	FD IV	Comparable FD OLS	FD IV
Annual change in per capita income	-0.2575*** [-3.58]	-0.2602*** [-3.92]	-0.0837 [-0.99]	-0.2030 [-1.43]
Annual change in Gini coefficient	1.0187*** [4.49]	0.9924*** [4.20]	0.2574 [1.01]	0.0013 [0.00]
PRSP adoption	-0.4636 [-0.69]	-0.6402 [-0.75]	-1.9713*** [-2.85]	1.6036 [0.70]
Period 4 dummy	1.1390 [1.56]	1.0578 [1.44]	-0.6375 [-1.19]	-0.3756 [-0.62]
R^2	0.3985*** [7.40]	0.4047*** [7.07]	0.1797** [2.67]	0.000 [0.35]
Observations	51	50	85	82

Source: Authors' calculations.

As a further exercise, the dollar-a-day base FD regression was re-estimated with two PRSP-interacted variables (in place of the adoption dummy) to determine the channel through which the estimated poverty reduction gains are operating.

The results, which are given in Table 5, confirm (as in the statistical test results) that poverty reduction is being secured through higher levels of growth alone, with the interacted per capita income variable significant (at the ten percent level) and right signed (column 2). The non-interacted Gini coefficient variable is again correctly signed and close to being significant at the ten percent level, whereas its interacted counterpart is not significant. This further underlines the

[16] Following standard practice, evidence of endogeneity was examined using an auxiliary regression testing procedure (Wooldridge, 2003: 532-533). This involves including the residuals from the first stage within the structural model as an additional independent variable, and testing the significance of the respective coefficient. These tests were statistically significant in the case of the dollar-a-day data (at the five percent level) but not the national data. Full details are available from the authors.

apparent lack of any distributional gains from PRSP adoption. The regression is also significant at the ten percent level and the R^2, although low at 0.13, has a similar value to the base regressions.

The FD results cast considerable doubt on the adoption benefit found in the statistical tests. While there remains an indication of superior PRSP performance in the base regressions, this is restricted to the dollar-a-day panel and is not maintained in the IV results. However, it might still be argued that the IV regressions, which are based on a highly restricted sample, are evidentially weaker. In addition, on the basis of the interacted results, they also show (if the presence of some effect is

Table 5: Results of second FD dollar-a-day data regression

1	2
FD of variable [t and F Statistics] (Dependent variable = headcount ratio)	FD OLS
Annual change in per capita income (PCY)	-0.0086 [-0.20]
Annual change in Gini coefficient	0.3963 [+1.33]
PRSP status interacted with PCY	-0.2689* [-1.87]
PRSP status interacted with Gini	-0.3290 [-0.69]
Period 3 dummy	+0.2943 [0.50]
Period 4 dummy	-0.1770 [-0.27]
R^2	0.1283* [1.88]
Observations	129

Source: Authors' calculations

accepted) that any poverty reduction gains are operating only through the growth channel. Thus there is no evidence of a distributional improvement in PRSP adopting versus non-adopting countries.

Two specific qualifications are worth making. Firstly, FD estimators do have certain measurement disadvantages; differencing tends to reduce variation, and the process cuts the available sample size. Secondly, these difficulties are made worse in this context by the parsing method needed to ensure a balanced panel, and this leads to some mismatching and interpolation of missing values. In all, these reduce the efficiency and statistical strength of the estimates, and hence also the reliability of any findings.

Fixed Effects (FE) estimation offers an alternative econometric treatment for the analysis of panel data, which deals with the challenges posed by unobserved heterogeneity but, additionally, can be more efficient than the FD approach in weaker data environments. The FE estimation process does not require a balanced panel, as it pools the observations, and allows for country-level effects by

effectively employing an intercept term for each record. As a result, sample sizes are substantially larger. In this situation, the number of observations within the national panel rises to 175, and within the dollar-a-day panel, to 256. However, under this approach the standard relation is specified in levels, and this departs from a direct replication of the core hypotheses being tested.

FE estimators can be implemented a number of ways, but here, a time de-meaning process is adopted, where the group mean for each country is deducted from each country observation. As such, the modelled variation is confined to that around country means. Again, the model is regressed on both the national and dollar-a-day datasets and employs a similar two-year lead-in binary variable to measure the presence of any treatment effect. As the full 12 years of data are pooled, this requires the use of 11 time variables (year 2 to year 12). As in the FD regressions, the base model was regressed using the full sample and a restricted version employing only consumption-based survey data. The precise specification is as equation 7, the double dot accent signifying the de-meaned value of the variables:

$$\ddot{P0} = \beta_1 \ddot{PCY} + \beta_2 \ddot{G} + \beta_3 \ddot{PRS} + \beta_{4\ldots14} \ddot{Yr}_{2\ldots12} + u \qquad [7]$$

In line with standard econometric practice, the models were also estimated using a Random Effects (RE) procedure. This approach, in contrast to FE, assumes the unobserved heterogeneity is variable and, therefore, deducts a weighted proportion of the group mean from each observation. RE estimation offers efficiency gains, but can threaten the consistency of the estimates. A Hausman specification test was performed to test for any efficiency improvements – in the event this found against any improvement and the FE results alone were retained.[17] As with the FD models, efforts were made to deal with possible endogeneity biases arising from the non-random nature of PRSP adoption, through the use of IV methods. However, although IFI indebtedness variables were again specified and tested (using a range of lags), it was not possible to find a significantly robust instrument. Therefore, we were unable to provide IV estimates for the FE models.[18]

Summary results for the FE regressions are given in Table 6 for the base and restricted OLS regressions. Overall, the results are positive for PRSP adoption and, notably, some significant benefit is evident in both panels. In the base OLS regressions, PRSP adoption is significant at the ten percent level (with a sizeable negative coefficient of 4.3) in the national data (column 2a); and significant at the one percent level (with a similar coefficient) in the dollar-a-day panel (column 3a). Within restricted sample regressions, the effect disappears in the national data (column 2b), whereas it actually strengthens in the dollar-a-day panel (column 3b). These findings suggest that PRSPs are having a substantial impact on poverty levels, although, again, the effect is more secure in the dollar-a-day panel. Further iterations in which the lead-in time was varied showed the apparent performance gain to be fragile in the national data, but strong in the dollar-a-day data. Unreported analysis shows the PRSP effect in the national data is only significant when a two-year interval is assumed.

In terms of the other variables, income is separately significant in the national data, albeit with a low coefficient of around -0.2, at the one percent level in the full, and at the five percent level in the

[17] For a full discussion of Fixed versus Random Effects, see Wooldrige (2006) pages 493-498.

[18] However, as the full results show, these were found to be inadequate instruments for PRS adoption. As a result, IV results are not provided for the FE models, and this is a major limitation.

Table 6: Base OLS Fixed Effects (FE) results

1	2a	2b	3a	3b
Variable (in levels) [t & F statistics] Dep variable = poverty rate	National data		Dollar-a-day data	
	FE OLS	FE OLS (Ex. income)	FE OLS	FE OLS (Ex. income)
Per capita income	-0.1852*** [-2.74]	-0.1631** [-2.49]	-0.0040 [-0.15]	-0.0096 [-0.24]
Gini coefficient	0.1124 [+0.38]	0.1375 [0.44]	0.1178 [0.88]	-0.013 [-0.08]
PRSP adoption	-4.2841* [-1.92]	-3.0432 [-1.25]	-4.2664*** [-2.89]	-4.9580*** [-2.78]
1997 dummy	-2.2126 [-0.63]	-0.7801 [-0.20]	-2.5301 [-1.81]	-4.5798 [-0.160]
1998 dummy	-2.1350 [-0.75]	-1.6177 [-0.53]	-0.6650 [-0.51]	-1.0470 [-0.46]
1999 dummy	0.5567 [0.14]	1.4529 [0.33]	-0.0369 [-0.02]	0.0564 [0.02]
2000 dummy	-0.0373 [-0.01]	0.9691 [0.29]	-1.3749 [-1.46]	-2.0161 [-1.11]
2001 dummy	-2.0082 [-0.75]	-1.8208 [-0.61]	-1.1781 [-0.85]	-1.8558 [-0.83]
2002 dummy	-1.9374 [-0.69]	-2.7370 [-0.89]	-0.6177 [-0.53]	-0.9953 [-0.50]
2003 dummy	-0.3626 [-0.11]	-1.4362 [-0.41]	0.4232 [0.29]	0.3168 [0.10]
2004 dummy	-0.6051 [-0.15]	-2.9503 [-0.72]	-3.6323** [-2.12]	-4.8225 [-1.78]
2005 dummy	-3.9578 [-1.14]	-5.651 [-1.41]	-2.4186* [-1.73]	-1.9489 [-0.75]
2006 dummy	-2.6908 [-0.71]	-2.6042 [-0.57]	-4.2340*** [-2.95]	-5.8336** [-2.31]
2007 dummy	7.6660* [1.80]	5.4047 [1.15]	-5.0344** [-2.07]	-6.6290** [-2.32]
Within R squared	0.3400*** [3.09]	0.3711*** [2.75]	0.2421*** [3.92]	0.2783*** [3.66]
Observations [groups]	175 [58]	150 [48]	256 [63]	171 [49]

Source: Authors' calculations.

restricted sample (columns 2a and 2b)); and some large time effects are evident in the dollar-a-day panel – with coefficients ranging from -0.2 for 2005 to -5.0 for 2007 (columns 3a and 3b). These suggest that poverty reduction accelerates from 2004 in the full sample. Although these effects are

replicated to some extent in the restricted sample, they are of varying significance and contradict the time pattern found in the FD results, which showed weaker poverty reduction in the final period. Both models are significant, but the explanatory power, given by the R^2 values, is higher for the national panel.

As with the FD analysis, attention was given to investigating the channel through which any PRSP impacts were operating. Again, this took the form of repeating the base regressions with two interaction terms in place of the binary treatment variable. The results are reported in Table 7, the national panel regression is given in column 2 and the dollar-a-day in column 3. The results for the national data reveal very little, with neither of the interacted variables found to be significant. Indeed, these findings somewhat undermine the validity of the PRSP-supporting result within the base regressions (column 2). The dollar-a-day results are, however, considerably more illuminating. The income interacted variable has both the right sign and is significant (at the five percent level); whereas the Gini interacted term is not (column 3). This again underlines that it is the growth channel through which any benefits are being felt. This matches the FD results and the findings of the statistical tests. The broad patterns revealed in the main regressions are also maintained in the results, with the income variable strongly significant in the national panel, and positive time effects again showing for the later years.

Taken together, the FE regressions go some way in rebalance the evidential position, reasserting the presence of a positive PRSP treatment effect. However, the position is again far from conclusive and three important qualifications are necessary. First, it remains the case that the evidence is largely confined to the dollar-a-day panel. Although there is a weakly significant result in the national data, it is exceptionally fragile, and this falls out in the restricted results and when the lead-in time is varied. Second, and importantly, our inability to provide a set of comparable IV regression results means it is not possible to reach a view on the potential for endogeneity bias within the FE results. This is especially problematic, given such biases are evident within the FD regression analyses. Third, as with the testing and FD results, it is apparent that the poverty reduction benefits associated with adoption are being brought about by better growth outcomes alone. Contrary to the policy objectives of the initiative, there no real evidence of more effective distributional management within PRSP countries.

Table 7: FE Regressions with PRSP interaction terms

1	2	3
Variable (in levels) [t & F statistics] Dependent variable = poverty rate	National	Dollar- a- day
	FE OLS	FE OLS
Per capita income	-0.2369*** [-2.94]	0.0234 [1.00]
Gini coefficient	0.1495 [0.49]	0.1120 [0.84]
PRSP interacted with income	0.0464 [0.83]	-0.0660** [-2.01]
PRSP interacted with Gini	-0.2211 [-1.24]	0.1156 [1.24]
1997 dummy	-2.211 [-0.61]	-2.3973* [-1.71]
1998 dummy	-2.0545 [-0.70]	-0.8045 [-0.62]
1999 dummy	0.6570 [0.16]	-0.0204 [-0.01]
2000 dummy	0.4048 [0.13]	-1.6795* [-1.74]
2001 dummy	-1.3525 [-0.48]	-1.4741 [-1.03]
2002 dummy	-1.4431 [-0.49]	-1.3602 [-1.15]
2003 dummy	-0.4569 [-0.14]	-0.0186 [-0.01]
2004 dummy	-0.0652 [-0.02]	-4.4385** [-2.53]
2005 dummy	-3.6894 [-1.02]	-3.4061** [-2.42]
2006 dummy	-2.3138 [-0.61]	-4.900*** [-3.37]
2007 dummy	7.1434* [1.69]	-6.0921** [-2.87]
Within R squared	0.3357*** [3.08]	0.2475*** [3.69]
Observations [groups]	175 [58]	256 [63]

Source: Authors' calculations.

6. Conclusions

The focus of this paper is on the outcome performance at the aggregate level of PRSP adoption. In addition, it provides an investigation of the character and neutrality of policy choices under PRSP arrangements.

The cross-sectional analyses sought to identify a positive treatment effect associated with PRSPs, defined in terms of higher levels of poverty reduction, and/or its proximate drivers of better economic growth and greater distributional equity. The weakness of the data record required the construction of two data sets and the application of a series of progressively more complex appraisal techniques. However, reaching definitive conclusions on the basis of the results which emerged is far from straightforward. The decision problem is summarised in the matrix given in Table 8. This subjectively classifies the strength of any apparent PRSP benefit at each stage of the analysis, as being either non-existent, moderate, or strong. The difficulty of reaching firm conclusions is apparent. Nevertheless, it seems possible for the evidence to support four substantial claims.

First, and most significantly, there is evidence of performance gains associated with PRSP adoption. This is in relation both to poverty reduction and growth outcomes. Furthermore, although the supportive evidence varies, some benefit is evident at each level of the analysis, and albeit more tentatively, within both datasets.

Table 8: Summary of analytical results

Type of analysis	**National dataset**	**Dollar-a-day dataset**
Statistical tests		
- Full sample	Weak – growth impact only	Strong – poverty and growth impacts
- Restricted sample		Moderate – weak poverty impact
First Differences Estimation		
- OLS	None	Strong – poverty impact (via growth)
- OLS consumption data only	None	Strong – poverty impact (via growth)
- IV	None	None
Fixed Effects Estimation		
- OLS	Moderate – poverty impact	Strong – poverty impact (via growth)
- OLS consumption data only	None	Strong – poverty impact (via growth)

Source: Review of authors' calculations.

However, this conclusion cannot be made without noting several specific qualifications, and these are in addition to wider evidential issues. Foremost, these impacts are largely confined to the dollar-a-day dataset. Where a benefit is apparent within the national panel, it relates more to growth outcomes, and is fragile to changes in sample and method. Yet, problematically, PRSPs simply do not target dollar-a-day poverty rates and thus, if we accept this evidence, we are doing so on the

basis of a policy goal which these national frameworks never sought to influence. In addition, the evidence of superior performance does weaken as the methods become more sophisticated. This is apparent at all stages of the analysis, but especially within the FD regression results, where the IV results fail to uncover any the treatment effect.

Second, with respect to policy questions, it can be concluded that there is little indication that PRSPs represent merely the latest form of IFI-sponsored structural adjustment. This conclusion is supported both by the set of statistical tests of stabilisation outcomes and, in turn, by the findings in relation to growth and poverty reduction. This does not necessarily contradict the suggestion that PRSPs place (too) much emphasis on macroeconomic stability, but it does suggest that the outcomes have not been as disinflationary as has been claimed by the most radical critics.

Third, it can be concluded that the higher levels of poverty reduction tend to be driven by enhanced growth outcomes alone. The interacted variable regressions and the initial testing make clear the primacy of growth. This indicates a lack of a pro-poor orientation within PRSP arrangements relative to any strong notion of pro-poor growth. This finding runs counter to both the spirit and operational claims of the initiative. Moreover, in the light of this, and the failure to find aggregate evidence of stabilisation biases, it is possible to construe PRSPs as reshaped, but narrowly-based, growth strategies. Such a characterisation fits well with the consensus thinking which underpins the PRSP policy template. This consensus neglects issues of distribution and favours the targeting of near-term growth, through a series of liberalisation measures.

Fourth, it is also clear that the evidential basis of the analysis is rather compromised. Two distinct sets of reservations need to be recognised. The first set relate to the quality of the underlying data, including the small sample sizes and issues of the comparability and consistency of the poverty data (within both panels). These concerns generally cast doubt on our ability to identify the presence of a genuine treatment effect. But there is also a more serious dimension – one of self-selection bias; only 29 of 68 adopting countries possessed data of sufficient quality to be included in the panel data. Yet it is likely that the availability of reliable data is itself a proxy for governmental and institutional capacity. If true, the selected treatment group will tend to over-represent those countries which are better equipped to devise and implement poverty reduction policies. In turn, the counterfactual comparisons employed above will systematically tend to overstate PRSP performance.

The second group of concerns relates to the use of aggregative techniques for policy analysis. By their nature, these tools can reveal little about the complex causes and dynamics which shape policy driven outcomes within countries. Arguably, PRSPs are especially difficult to examine through these approaches, due to their claimed policy neutrality and doubts over the extent to which they represent real world policy frameworks, as opposed to merely pro-forma commitments. In order to establish causation, what can be termed identification and attribution problems would need to be resolved. The identification problem relates to the extent that PRSPs can genuinely be said to comprise a set of policy choices, or at least, to represent a defined policy stance. The second issue of attribution relates to the extent of implementation and is potentially a more difficult issue to resolve, since it would require that this be reflected in the dataset. The analysis currently assumes that possession of a PRSP for two years or more implies implementation of the strategy. Attribution was addressed in the earlier policy-based lending literature by means of slippage variables identified by the IFIs during the course of an arrangement. Yet such an option is not

available for PRSPs. These evidential objections do not amount to an outright rejection of aggregate-level appraisal techniques but, rather, suggest that the findings above at the very least require triangulation. The key issue is the need to establish the causal connections between adoption, policy choices and outcomes. Therefore, country-based evaluations may prove to be a more fruitful means of assessing PRSP performance.

References

Booth, D. (ed.) (2003). *Fighting Poverty in Africa: Are PRSPs Making a Difference?* London: ODI.

Bourguignon, F. (2004). 'The poverty-growth-inequality-triangle'. Paper presented to the Indian Council for Research on Economic Relations, New Delhi, February.

Bourguignon, F. (2002). 'The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods'. In Theo S. Eicher and Stephen J. Turnovski (eds.), *Inequality and Growth Theory and Policy Implications* (Chapter 3). Cambridge, MA: MIT Press.

Chen, S. and Ravallion, M. (2008). 'The developing world is poorer than we thought, but no less successful in the fight against poverty'. *World Bank Policy Working Paper Series*, No. 4703.

Christiansen, K. and Holland, I. (2003). 'The PRSP Initiative: multilateral policy change and the role of research'. London: *ODI Working Paper Series*, No. 216.

Datt, G. and Ravallion, M. (1992). 'Growth and redistribution components of changes in poverty measures: a decomposition with application to Brazil and India in the 1980s'. *Journal of Development Economics*, 38, 275-295.

Independent Evaluations Group (formerly the Operations Evaluation Department) of the World Bank (IEG) (2004). *The Poverty Reduction Strategy Initiative: a Review of the World Bank's Operations through 2003*. Washington, DC: World Bank

Independent Evaluations Office of the IMF (IEO) (2004). *Evaluation of the IMF's Role in Poverty Reduction Strategy Papers and the Poverty Reduction and Growth Facility*. Washington, DC: IMF.

Kakwani, N. (1997). 'On measuring growth and inequity components of changes in poverty, with application to Thailand'. *Discussion Paper No. 97/17.* Sydney: University of New South Wales (UNSW) School of Economics.

Marshall, R. (2010). *An Appraisal of the Poverty Reduction Strategy Initiative*. Unpublished PhD thesis, University of Manchester.

Ravallion, M. (1994). *Poverty Comparisons*. London: Routledge.

Ravallion, M. and Chen, S. (1997). 'What can new survey data tell us about recent changes in distribution and poverty?' *The World Bank Economic Review*, 11, 357-382.

Reddy, S. and Pogge, P. (2005). 'How *not* to count the poor'. *Social Analysis Working Papers.* Online resource available: http://www.socialanalysis.org/ (accessed 26 February 2011).

Wooldridge, J. (2003). *Introductory Econometrics: a Modern Approach*. Mason, OH: Thomson Learning/ South Western College Publishing.

The University of Manchester
Brooks World Poverty Institute

MANCHESTER 1824

Executive Director
Professor David Hulme

Research Directors
Professor Armando Barrientos
Professor Rorden Wilkinson

The Brooks World Poverty Institute (BWPI) creates and shares knowledge to help end global poverty.

BWPI is multidisciplinary, researching poverty in both the rich and poor worlds.

Our aim is to better understand why people are poor, what keeps them trapped in poverty and how they can be helped - drawing upon the very best international practice in research and policy making.

The Brooks World Poverty Institute is chaired by Nobel Laureate, Professor Joseph E. Stiglitz.

Contact:

Brooks World Poverty Institute
The University of Manchester
Humanities Bridgeford Street Building
Oxford Road
Manchester
M13 9PL
United Kingdom

Email: bwpi@manchester.ac.uk

www.manchester.ac.uk/bwpi

www.manchester.ac.uk/bwpi