

How valid are synthetic panel estimates of poverty dynamics?

Nicolas Héroult

Melbourne Institute of Applied Economic and Social Research, University of Melbourne

Email: nherault@unimelb.edu.au

Stephen P. Jenkins

London School of Economics, ISER (Essex), University of Melbourne, and IZA

Email: s.jenkins@lse.ac.uk

30 January 2019

Abstract

A growing literature uses repeated cross-section surveys to derive ‘synthetic panel’ data estimates of poverty dynamics statistics. It builds on the pioneering study by Dang, Lanjouw, Luoto, and McKenzie (‘DLLM’, *Journal of Development Economics*, 2014) providing bounds estimates and the innovative refinement proposed by Dang and Lanjouw (‘DL’, *World Bank Policy Research Working Paper 6504*, 2013) providing point estimates of the statistics of interest. We provide new evidence about the accuracy of synthetic panel estimates relative to benchmarks based on estimates derived from genuine household panel data, employing high quality data from Australia and Britain, while also examining the sensitivity of results to a number of analytical choices. For these two high-income countries we show that DL-method point estimates are distinctly less accurate than estimates derived in earlier validity studies, all of which focus on low- and middle-income countries. We also demonstrate that estimate validity depends on choices such as the age of the household head (defining the sample), the poverty line level, and the years analyzed. DLLM parametric bounds estimates virtually always include the true panel estimates, though the bounds can be wide.

Keywords: synthetic panel, pseudo panel, poverty dynamics, poverty entry, poverty exit, BHPS, HILDA

JEL Classifications: I32, D31, C52

1. Introduction

There is a growing literature that employs repeated cross-section surveys to derive ‘synthetic panel’ data estimates of poverty dynamics statistics building on the pioneering study by Dang, Lanjouw, Luoto, and McKenzie (2014, hereafter ‘DLLM’) providing bounds estimates and the innovative refinement proposed by Dang and Lanjouw (2013, hereafter ‘DL’) providing point estimates of the statistics of interest. All but one of the applications to date of these methods, of which there are many, have been to middle- and low-income countries. This paper provides new evidence about the validity of synthetic panel estimates, employing high-quality and long-running household panel data from two rich countries, the British Household Panel Survey (BHPS) and the Households, Income, and Labour Dynamics in Australia survey (HILDA).

DLLM state clearly the reasons for employing a synthetic panel approach:

Genuine panel data are still rare in the developing world, and when they are available, the samples are often relatively small, with limited or infrequent duration, and in some cases, occur with significant attrition. This has limited the feasibility of constructing even relatively simple descriptions of transitions in and out of poverty for most countries. Yet policymakers and researchers do care about such transitions, and most countries do field repeated cross-sectional surveys of income or consumption on a reasonably regular basis.

(DLLM: 123–4.)

DLLM focus on statistics summarising poverty status in two years, e.g. the joint probability of being poor in one year and poor in the second year, and develop methods that provide both non-parametric and parametric bounds on this probability and each of the three other joint poverty probabilities. They assess the validity of their approach using panel data: synthetic estimates, derived by treating the panel data as two independent cross-sections, are compared with the ‘true’ estimates derived from the longitudinal data per se (more about this below).

DLLM conclude, using data for Indonesia and Vietnam, that ‘the bounds can be narrow enough in practice to make the estimates useful’ (DLLM: 124). Cruces et al. (2015) assess the DLLM bounds method using data from Chile, Nicaragua, and Peru, incorporating extensive examination of the sensitivity of their results to a number of analytical choices about definitions, and conclude that ‘the methodology performs reasonably well’ (2015: 163). Perez (2016) also assesses the DLLM bounds method but using Mexican data.

DL refine the DLLM method and derive point estimates of poverty probabilities rather than bounds. Their empirical analysis, based on data for Bosnia-Herzegovina, Laos, Peru, Vietnam, and the USA, supports the validity of their refined method:

[W]e show that our estimates are quite accurate ... We find that estimation results are good not only for the general population but for smaller population groups as well, and are associated with much tighter confidence intervals than even direct, panel-data based estimates in those settings where the sample sizes for the cross sections are large enough. (DL: 36.)

The only other validation study of the DL method to date is by Garcés Urzainqui (2017), who concludes from his analysis of data for Thailand that ‘the general patterns of mobility described by synthetic panel estimates are well in line with the true dynamics’ (2017: 35).

DLLM’s and DL’s synthetic panel methods have also been used to derive poverty dynamics estimates for a large number of middle- and low-income countries without any accompanying validation exercise. Three studies have applied the DLLM bounds method: Ferreira et al. (2013) study 18 Latin American countries; Rama et al. (2014), India and Bangladesh; and Perez (2016) using a different Mexican data set from that used in his validation exercise. There are (at least) seven studies applying the DL point estimates method: Dang et al. (2017) to Senegal; Dang and Ianchovichina (2018) to 6 Middle-Eastern and North African countries; Rigolini et al. (2016), to 17 Latin American countries; Dang and Dabalén (2018), to 20 Sub-Saharan African countries; and Dang and Lanjouw (2018), to India. OECD (2015: Appendix 5.A1) apply the DL method to measure transitions into and out of low-pay status for individual workers in 10 emerging economies. OECD (2018) apply the DL method extensively in their cross-country analysis of movements into and out of the poorest fifth and the richest fifth of the income distribution.

In sum, there are now many applications of synthetic panel methods to a large number of countries in the developing world, and some research suggests that they produce valid estimates. However, there is a need for further validation studies of synthetic panel methods of estimating poverty dynamics statistics. This paper meets that need and makes a number of contributions in addition. We focus our analysis on the DLLM and DL methods because they have been used in virtually all applications to date. A potential competitor method is that proposed by Bourguignon and Moreno (2015) which uses a different income model from DL’s. However, Bourguignon and Moreno’s approach is harder to implement, and has only been applied in one other study so far (Garcés Urzainqui 2017).

Although there is a large number of applications of the DL method, there are few validation studies. We assess the DL method in detail, also incorporating sensitivity checks to a number of analytical choices about definitions analogous to the way in which DLLM and Cruces et al. (2015) assessed the performance of the DLLM method. We also provide DLLM parametric bounds estimates.

We add substantially to analysis of the DL and DLLM methods in rich country contexts with our study of Australia and Britain; DL's US case study is the only previous rich country application of the synthetic panel approach. Our paper is the first to consider how variations in the age range of the household head used to define analysis samples affects results for a given country. (Age ranges vary across earlier studies but not within them.) We also consider the impact of using different definitions of the cohorts used to derive the parameter ρ which is a fundamental ingredient of the DL method (explained below). Only Garcés Urzainqui (2017) has done this before.

Our research is also distinctive because we examine the sensitivity of the DL method to the choice of the poverty line for the first time. DLLM and Cruces et al. 2015 examined the sensitivity of the DLLM bounds approach (not the DL one). This turns out to be important. In addition, we derive poverty dynamics estimates for subgroups of individuals defined by age (0–17, 18–59, and 60+ years).

Further distinctive features of our work are as follows. For all sets of analytical choices about definitions, we derive estimates of poverty exit and entry rates, i.e. estimates of two conditional probabilities, in addition to estimates of the four joint probabilities that have been the focus of previous work. In this respect, we are taking up the challenge of Fields and Viollaz (2013: 20) who argue that it is these conditional probabilities that are more relevant to the essence of 'poverty dynamics' and who contend that the DLLM method estimates them less accurately. DL provide conditional probability estimates, noting that they are 'slightly less accurate' than the joint probability estimates (p. 33). We provide new and more extensive evidence about whether estimates of conditional poverty probabilities are more or less accurately estimated than joint probabilities.

In addition, because the BHPS and HILDA are much longer-running household panels than those for any developing country – we use data collected annually over 18 years for the BHPS, and over 15 years for HILDA – we can provide a detailed assessment of the extent to which the accuracy of synthetic panel estimates of poverty dynamics statistics vary according to the year or period studied. This turns out to be important.

Finally, we consider the benchmarks used to assess the accuracy of the synthetic panel estimates. As we discuss below, and has not been pointed out before, the ‘true’ benchmarks employed by DL differ from those that are typically used in ‘standard’ panel data approaches to poverty dynamics. We show how the benchmark estimates shift if one changes the following rule and consider the implications for assessments of accuracy of synthetic panel estimates.

Although the application of DLLM and DL methods has focused on developing countries, there is value in assessing their validity in rich country contexts even though panel data are more common. For example, there are long-standing concerns about attrition in the longitudinal data components of the EU’s Statistics on Income and Living Conditions (EU-SILC), i.e. the sources used to calculate the EU’s measure of persistent poverty. Among the countries using household panel surveys to collect data about income and poverty annually over a four-year period, there is substantial loss to follow-up. For example, Jenkins and Van Kerm (2017: Figure 22.1) show that around one half of the countries using surveys have four-year (2008–2011) retention rates of less than 70% with the smallest rate just over 40% (UK). By comparison, the four-year retention rates following the first waves of high-quality panels such as the BHPS and HILDA are nearly 80% (Watson and Wooden 2011, Figure 3).

The rest of our paper unfolds as follows. In Section 2, we review how the DLLM and DL methods work, and point out the key analytical choices that are required to implement them. In Section 3, we describe our HILDA and BHPS data and explain our various definitional choices. We report our empirical results in Sections 4 and 5. Section 4 examines the accuracy of estimation of the cross-year correlation parameter (ρ , i.e. ‘DL rho’) that underpins the DL method, and which is derived using pseudo-panel methods. We show that, depending on the definition of the cohorts and the age range of the household head used to define the analysis sample, DL-method estimates of ρ can vary substantially depending on the time period considered and also be very different from the ‘true’ panel data benchmark estimates.

In Section 5, we report our assessments of the validity of DL method estimates of joint and conditional poverty statistics, focusing on a ‘leading case’ set of choices relating to the definition of cohorts, sample selection (the age range of household heads), and the poverty line. In Section 6, we document how assessments change as we vary, in turn, choices about the ‘true’ panel benchmark (the following rule issue), the age range of household heads, and the poverty line. Finally, we look at estimates of poverty dynamics statistics for

three age groups of individuals (aged 0–17, 18–59, and 60+ years). Our conclusions are in Section 7.

Overall, we show for Australia and Britain that DL-method point estimates of poverty dynamics statistics are distinctly less accurate than estimates derived in earlier validity studies, all of which focus on low- and middle-income countries. We also demonstrate that estimate validity depends on choices such as the age of the household head (defining the sample), the poverty line level, and the years analyzed. DLLM parametric bounds estimates virtually always include the true panel estimates, though the bounds can be wide.

For brevity, we report only a selection of results in the main text. Appendices A (HILDA) and B (BHPS) in the Supplementary Material report the estimates of the DL(LM) income model regressions, the means of the income predictors, as well as visual checks of the bivariate normality assumptions, year by year. We also provide a full set of estimates of all poverty dynamics statistics for each of the 28 different combinations of definitions we use. Estimates of the cohort regressions used to derive DL rho (see Section 2) are available from the authors on request.

2. How do the DLLM and DL methods work?

With cross-sectional survey data for a pair of years (Year 1, Year 2), one has information about the marginal distributions of income in each year. (The outcome variable might be consumption rather than income; we refer to the latter.) Clearly, there is no information about the joint distribution of income in the two years, nor thence information about poverty dynamics. The DLLM and DL methods work by using a model and associated assumptions to fill in the missing longitudinal information. In this section, we review the key elements of the two methods, drawing heavily on the original expositions.

The first step, common to DLLM and DL methods, is an *income model* for each of Year 1 and Year 2. Suppose that income y_{it} for household head i in year t is described by

$$\log(y_{i1}) = \beta_1'x_{i1} + \varepsilon_{i1} \quad (1)$$

$$\log(y_{j2}) = \beta_2'x_{j2} + \varepsilon_{j2} \quad (2)$$

where x_{i1} and x_{j2} are vectors of time-invariant predictors in Years 1 and 2.

The Year 1 income for each household head j observed in Year 2 is unobserved but it can be predicted using model estimates and two auxiliary assumptions. Ordinary least squares

regression applied to each of (1) and (2) yields parameter estimates $\hat{\beta}_1, \hat{\beta}_2$ (the regression coefficients), $\hat{\sigma}_1, \hat{\sigma}_2$ (the variances of the residuals in each year), and also of the residuals $\hat{\epsilon}_{i1}, \hat{\epsilon}_{j2}$ for each i, j . DLLM's Assumption 1 is that the population sampled by the two cross-section surveys is the same in Year 1 and Year 2 and so, for instance, the distributions of the time-invariant regressors are the same. DLLM's Assumption 2 states that the residual errors ϵ_{i1} and ϵ_{i2} are 'positive quadrant dependent', a property that includes being positively correlated.

DLLM show that with these assumptions one can derive *nonparametric bounds* on the four joint poverty probabilities. The Upper Bound scenario (with maximum income mobility) arises when the error terms in the Year 1 and 2 income equations are independent. The Year 1 income distribution can then be predicted as follows. For every observation j in the Year 2 sample, take a random draw with replacement from the empirical distribution of residuals for Year 1 (with mean 0, s.d. $\hat{\sigma}_1$) and predict the outcome in Year 1 using the expression $\log(\widehat{y_{j1}^U}) = \hat{\beta}_1'x_{j2} + \check{\epsilon}_{j1}$, where $\check{\epsilon}_{j1}$ is the residual imputed to each j . We now have synthetic panel data from which poverty dynamics statistics can be calculated. Although the observation unit is the household head, application of appropriate survey weights (that also account for the number of individuals in each head's household) provides estimates referring to the population. To counter the variability introduced by the stochastic nature of the imputations, one repeats the random-draw-and-calculation step R times and averages the resulting estimates. We find that setting $R = 50$ is sufficient.

DLLM's Lower Bound scenario arises when the residual errors are perfectly correlated across Year 1 and Year 2. In this case, the Year 1 income prediction for each j from the Year 2 sample is $\log(\widehat{y_{j1}^L}) = \hat{\beta}_1'x_{j2} + \gamma\hat{\epsilon}_{j2}$, where scalar γ is chosen to ensure the standard deviation of the imputed Year 1 residuals distribution equals $\hat{\sigma}_1$. Again, we now have synthetic panel data and can calculate the poverty dynamics statistics of interest, again using weights to derive population-level estimates.

DLLM also proposed a *parametric bounds* approach in order to narrow the bounds, arguing that non-parametric bounds for the poverty dynamics statistics may be wide and hence not particularly useful in practice. DLLM's key additional assumption is that the distribution of errors in (1) and (2) is bivariate normally distributed. This means that all the poverty dynamics statistics can be calculated if one has an estimate of the correlation of the errors, ρ . The bivariate distribution of income is fully characterised by ρ , the estimates of the standard deviations of the cross-section marginal distributions (σ_1, σ_2), and the other income

model parameters. For example, the probability of being poor in Year 1 and also poor in Year 2 is

$$\text{Prob}(y_{j1} < z_1 \text{ and } y_{j2} < z_2) = \Phi\left(\frac{\log(z_1) - \beta_1' x_{j1}}{\sigma_1}, \frac{\log(z_2) - \beta_2' x_{j2}}{\sigma_2}, \rho\right) \quad (3)$$

where $\Phi(\cdot)$ is the bivariate normal cumulative distribution function, and z_1 and z_2 are the poverty lines for Year 1 and Year 2. There are analogous expressions for the other three joint poverty dynamics probabilities (DLLM, section 4).

DLLM suggest choosing bounds for ρ by drawing on information from longitudinal surveys for the same country or other similar countries. In their applications, they use ρ values of (0.2, 0.8) and (0.3, 0.7).

DL build on DLLM's parametric bounds approach, innovatively showing that one can derive a point estimate for ρ (and thence for each of the poverty statistics of interest) from the data already to hand rather than by relying on auxiliary estimates from other surveys to provide bounds. (There are also a number of other extensions, including applications to income dynamics over more than two years, and to mobility between more than two income classes, but we do not examine these aspects here.) DL's key insight is that pseudo-panel methods utilising panels based on cell mean data about cohorts can be deployed. Cohorts are defined by grouping together individuals sharing the same (or similar) age and time-invariant characteristics (e.g. sex, ethnic background). Important references that DLLM draw on include Deaton (1985), Moffitt (1993), and Verbeek (2008).

DL show first, in their Proposition 1, that an approximate estimate of the correlation between $\log(y_1)$ and $\log(y_2)$, $\rho_{y_1 y_2}$, can be derived from a cohort-level regression of Year 1 on Year 2 cohort mean incomes. (Our derivation follows DL Appendix 1 eq. 1.6.) This is an Instrumental Variables estimator and hence (as DL point out) reliant on a number of assumptions, some of which untestable and have to be maintained. Intuitively, one needs substantial income variation across cohorts so that cohort means are sufficiently predictive and also sufficiently large cohort sizes to ensure sufficient precision.

Second, DL show in their Proposition 2 that the all-important cross-year correlation of residuals, ρ , can be derived from $\rho_{y_1 y_2}$ and other information already to hand from the income models:

$$\rho = \frac{\rho_{y_1 y_2} \sqrt{\text{var}(\log(y_{j1})) \text{var}(\log(y_{j2}))} - \beta_1' \text{var}(x_j) \beta_2}{\sigma_1 \sigma_2}. \quad (4)$$

This is the 'DL rho' estimate that we report below. DL also show that another estimate of ρ

can be derived (Corollary 2.1), though they state (p. 13) that it typically provides very similar estimates. We find this too, and so do not report this other estimate.

Given the estimate of ρ , estimates of poverty dynamics statistics can be derived using the same approach as set out for the DLLM parametric bounds approach: cf. (3) above, and analogous expressions for the other joint probabilities as well as poverty entry and exit rates.

DL's theoretical results provide little direct guidance to analysts about how to define the cohorts in practice. Their own empirical analysis uses cohorts defined in terms of age only (DL: 30), with samples restricted to households for which the household head in Year 1 is aged 25–55 years, and they do not consider the practical implications of alternative definitions. In contrast, Garcés Urzainqui (2017, especially section 5.3) discusses theoretical and practical issues concerning cohort definitions extensively.

We conclude from this research that there is no single definition of cohorts that is clearly the 'best' for empirical work and, correspondingly, there is substantial scope for analysts to choose among a relatively large number of potential definitions. Put differently, the choice of cohort definition, and the closely related sample selection decision concerning the age range of the household head, is a potential source of sensitivity for synthetic panel estimates of poverty dynamics statistics that needs to be investigated because it is fundamental to the DL method. We provide this analysis. Further information about our analytical choices is provided in the next section.

We assess *accuracy* ('validity') as previous researchers do by examining how close each of the synthetic panel estimates is to the corresponding estimate derived from genuine panel data. The latter is treated as the 'true' estimate, with the quotation marks used to indicate that the benchmark is itself an estimate. We follow DLLM and DL by counting a synthetic panel estimate as sufficiently accurate if it lies with a 95% confidence interval of the 'true' estimate. DL also use a tighter criterion (within one standard deviation of the 'true' estimate) and also some coverage criteria (DL: 31), but we find these unnecessary in our applications because validity is not achieved for many estimates using the looser criterion (see later). Other approaches to assessing accuracy are possible, e.g. one might compare the 'true' point estimate to the 95% confidence interval of the synthetic panel estimate, or look at confidence interval overlap for both estimates. However, because our focus is the validity of the DLLM and DL methods as usually implemented, we use their main approach to assessing validity.

3. HILDA and BHPS: data and definitions

We use household panel data from waves 1–15 of HILDA (covering 2001–2015) and from waves 1–18 of the BHPS (covering from 1991 to 2008, its final year). The surveys share a common design: the original respondents are a sample of the private household population of the country concerned and are re-interviewed annually. Both HILDA and the BHPS follow individuals from originally-sampled and split-off households, like the US Panel Study of Income Dynamics and the German Socio-Economic Panel. As in the SOEP (but not the PSID), HILDA and the BHPS interview all adults within a household. HILDA and the BHPS are widely renowned as very high-quality datasets. For example, attrition rates are low, around 5 percent or less each year after wave 1 (Watson and Wooden 2015). Both surveys provide weights to derive population-level estimates that account for non-response (including attrition). For an overview of HILDA and the BHPS, see Frick et al. (2007).

The measure of living standards is the same in HILDA and the BHPS, i.e. equivalised net household income. See Wilkins (2017: Chapter 3) for HILDA and Jenkins (2011, chapter 4) for the BHPS. Household net income is (a) total money income earned in the labour market from employment or self-employment, income from the capital market (e.g. stocks, shares, interest-bearing accounts, and other financial assets), cash transfers from the government, plus private transfers, from which is deducted (b) national and local income taxes and (in Britain) social insurance payments, and a small number of other deductions. Net income is adjusted to account for differences in household size and composition using the ‘Modified OECD’ equivalence scale. The income definition is consistent with the recommendations of international bodies such as the Canberra Group (2011), and is the definition used by the OECD and Eurostat to produce their inequality and poverty statistics. To include a very small number of zero values for income (between 7 and 26 per year in HILDA; between 1 and 2 in the BHPS), we follow DL (p. 6) and apply a modified Box-Cox transformation to observed incomes.

We set the poverty line at 60% of contemporary national median income in most of our analysis, but also consider 50% of contemporary national median income as an alternative. The 60% cut-off is used by UK official statistics and Eurostat to derive their ‘headline’ poverty statistics. Australia and the OECD have commonly used the 50% threshold. DL(LM) mostly use official poverty lines in their studies and so we are following them in this respect. An important difference is that our poverty lines are relative lines whereas the official

poverty lines for the countries that DL(LM) consider are absolute poverty lines typically derived with reference to calculations of minimum cost food and non-food budgets.

We undertake analysis using samples selected according to two definitions of the age of the household head: 25–55 years and 25–75 years. (These refer to the age in Year 1; the Year 2 age range is adjusted upwards accordingly.) This range encompasses those used in previous research and also allows us to check how the validity of synthetic panel estimates varies with the choice. The HILDA sample sizes are between 5,190 and 7,198 household heads per year for samples with heads aged 25–75 years, and between about 3,618 and 4,880 for the samples with heads aged 25–55. For the BHPS, the corresponding sample sizes are between 2,464 and 3,190 and 1,526 and 2,176, respectively.

DLLM state that they use the 25–55 age range because ‘analysis of poverty transitions among households headed by those younger than 25 or older than 55 or 60 is more difficult since at those ages households are often beginning to form, or starting to dissolve’ (DLLM: 114). DL state that their choice is ‘consistent with the literature on pseudo-panel data ... While this age range can be extended to include older people, it may be ill-advised to include those who are younger, at least since most household heads tend to be older than 25 in all the countries we look at’ (DL: 30). Cruces et al. (2015) used the range 25–65 years, ‘in order to avoid lifecycle effects which can invalidate the time invariance assumption’ (2015: 166). Garcés Urzainqui selects heads aged 25–70, remarking that ‘limiting the age of the household head is a standard procedure in this literature to restrict attention to stable households, avoiding the age segments most associated with household formation and household dissolution. ... I am more relaxed ... [Thai] households with older heads tend to be poorer so that strict age limits may lead to a distorted view on poverty dynamics’ (2017: 21–22).

There are two issues here. One is the need to ensure the time-invariance assumption is satisfied. The second concerns the treatment of household formation and dissolution, but this is more an issue concerning the definition of the benchmark ‘true’ panel estimates of poverty probabilities and the panel survey’s following rule (the issue we flagged in the Introduction). Relevant to both issues is the question of the age ranges in which household formation and dissolution is most prevalent. By comparison with developing countries, in rich countries, households with heads aged 55–75 may be more stable than those with a head aged 25–55 because divorce and partnering is more common among the latter group, and longevity is greater. In any case, there is a separate argument in favour of using as wide an age range as possible because this provides estimates with greater coverage of the population and includes groups of policy interest such as elderly people.

The predictors that we include in the Year 1 and 2 income models are much the same as those employed by DL (see their Appendix 2). More specifically, for HILDA, the predictors are the household head's sex, birth cohort (by half decade), education level (four categories), and country of birth (whether of Australian origin, or a migrant from English speaking country or non-English speaking country). For the BHPS, the predictors are the household head's sex, birth cohort (by half decade), education level (five categories), and whether of non-white ethnic origin. Both models also include interactions between education and sex and education and birth cohort. All predictors are time-invariant, as assumed by the DL(LM) method. In preliminary work, we also considered simpler income model specifications, without the educational level variables and associated interactions. This choice led to reduced goodness of fit (i.e., smaller adjusted- R^2) but made little difference to our estimates of poverty dynamics statistics and so we do not report these results.

We use a one-year gap between Year 1 and Year 2. 'One year' corresponds to within a month of the anniversary of the previous year's interview for the vast majority of households in HILDA (Summerfield et al. 2016, Tables 8.5, 8.6) and the BHPS (calculations by the authors). Thus HILDA provides us with 14 year-pairs of data (2001/2002 through to 2014/2015), and the BHPS 17 year-pairs of data (1991/1992 through to 2007/2008).

Use of the one-year gap enables us to examine the sensitivity of validity checks to the period considered much more extensively than in other studies. Using a one-year gap is also favourable to the DL(LM) method because it is more likely that the population sampled by Year 1 and Year 2 cross-sections is the same (cf. Assumption 2). We found no evidence of a substantial reduction in the accuracy of DL(LM) estimates of poverty dynamics when we considered five-year gaps in preliminary analysis.

Estimates of ρ are contingent on the definition of the cohorts used, and so to analyse sensitivity, we consider multiple definitions. For HILDA, the first two of the seven cohort definitions are based on year of birth, YOB(s) where s , the number of birth years covered, is 1 and 5. (DL use the YOB(1) definition.) The third and fourth definitions are based on sex interacted with each of YOB(5) and YOB(10). Definitions 5 to 7 are based on YOB(5), YOB(3) and YOB(10), each interacted with country of birth (Australia, English-speaking, or non-English-speaking country). For the BHPS, the six cohort definitions are specified similarly, with cross-country differences reflecting variable availability and cell size considerations. Four definitions are based on YOB(s) where s is 1, 3, 5, and 10. The other two definitions use sex interacted with each of YOB(5) and YOB(10).

Numbers of cohorts and cohort size for each definition and country are shown in Table 1 (HILDA) and Table 2 (BHPS) below. There are two panels in each table corresponding to whether the sample selected refers to heads aged 25–55 or 25–75. The greater the number of cohorts, the smaller the average cell size. For example, for HILDA, definition YOB(1) has 51 cohorts with average size 123; for definition SEX*YOB(10), there are 10 cohorts with average size 619.

For the validation analysis per se, our data sets are constructed in the same way as DLLM’s (see their pp. 116–117). Each two-year longitudinal sample is split randomly into two samples A and B. The ‘true’ panel estimates are derived using sample A’s data for Year 1 and Year 2. The two cross-sections are the Year 1 data for sample A and the Year 2 data for sample B. This ensures that no individuals appear in both cross-sections; if they did, it could contaminate the exercise by introducing spurious cross-unit correlation. Because the sample splitting is done randomly, we report results for each country-year-pair based on the averages of 50 splits.

4. Estimates of DL rho using pseudo-panel methods

We report estimates of ρ for each cohort definition and sample selection criterion, as well as information about numbers of cohorts and cohort size, in Table 1 (HILDA) and Table 2 (BHPS). We do not show the results for every Year 1–Year 2 pair here; instead, and in order to highlight differences arising from cohort definitions and sample selection, Tables 1 and 2 show averages across year-pairs (14 for HILDA, 17 for the BHPS). Variability in estimates by year-pair is addressed shortly.

<Tables 1 and 2 near here>

For assessing the accuracy of the DL method, the most important rows of Tables 1 and 2 are the ones labelled ‘panel rho’ and ‘DL rho’. The former is the ‘true’ correlation of the residuals from the Year 1 and 2 income regressions, derived from the longitudinal sample defined earlier. The (averaged) ‘true’ rho is almost 0.75 for Australia and slightly smaller for Britain (nearer 0.70), but hardly varies with the household head’s age range in both countries.

DL rho is the (averaged) estimate of ρ derived from the income-cohort correlations and the transformation shown in equation (4). For both Australia and Britain, there is substantial variation according to cohort definition, and many estimates differ substantially

from the ‘true’ panel rho. For the YOB(1) cohort definition and sample with head aged 25–55, the definitions closest to DL’s, the HILDA panel rho is more than five times larger than the DL rho (Table 1) and the BHPS one is more than three times larger (Table 2). Yet, at the same time, the HILDA panel rho differs by no more than 15 percent of DL rho in six of the 14 cohort-age-range combinations. For the BHPS, the corresponding number is six out of 12 combinations. Estimates tend to be better when the head’s age is 25–75, a result that also applies to both countries. Indeed, for this head age range and for cohort definition COB*YOB(5) for Australia and Sex*YOB(5) for Britain, the averaged DL rho is virtually the same as the ‘true’ benchmark.

Our finding that the quality of DL rho estimates is sensitive to the choice of cohort definition is worrying because researchers applying the DL method (and without longitudinal data) might choose the ‘wrong’ definitions and sample selection criterion, if only because of data constraints. This could lead to inaccurate estimates of ρ and thence of the poverty dynamics statistics of interest.

The averaged estimates in Table 1 do not show how estimates vary across year-pairs. For brevity, we document temporal variation for only four combinations of cohort definition and household head’s age range. (These encapsulate the range of results derived from all the combinations.) Figure 1, panel (a) for HILDA and panel (b) for the BHPS, shows the year-by-year estimates of the ‘true’ panel rho (with pointwise 95% confidence interval) as well as the estimates of DL rho. Also shown are estimates of DL’s cohort-mean correlation coefficient, ρ_{y_1, y_2} , labelled ‘simple cohort correlation’ in the Figures. These estimates represent the first step in deriving DL rho estimates before information from the income model is incorporated in a second step (see eq. 4). DL rho estimates track the cohort correlations closely but tend to be lower. As the cohort correlations are more often an overestimate than an underestimate of the ‘true’ panel rho, DL rho is a closer approximation to the ‘truth’ than the cohort correlation on average, though that is not the case for all year-pairs.

<Figure 1 near here>

For Australia, the ‘true’ panel rho increases slightly over the period. However, for only one combination (head aged 25–75, cohort definition COB*YOB(5)) of the three is the ‘true’ panel rho tracked well by DL rho and, even in this case, there are some noticeable differences between them in the very earliest and very latest year-pairs. For the other three combinations, differences are markedly larger, and it is clear the choice of household head’s

age range is the principal contributor to the differences between the estimates. In the two charts on the right-hand side of Figure 1(a), DL rho differs substantially from the ‘true’ panel estimate, especially in the second half of the period, and fluctuates substantially. In addition the DL rho estimate generally declines over time, contrary to the slight rise in the ‘true’ panel rho.

For Britain (Figure 1(b)), many of the same patterns are apparent. The main differences compared to Australia are, first, that there is slightly more variation over time in the estimates of the ‘true’ panel rho. Second, the differences between estimates of DL rho and the ‘true’ panel rho are not as large as those for Australia, even for the two combinations with household head aged 25–55. This suggests that BHPS synthetic panel estimates of poverty dynamics statistics are likely to be more accurate than their HILDA counterparts, and less sensitive to the choice of definitions and sample selection.

Why there are large changes in DL rho estimates over time is unclear. There were no changes in HILDA or BHPS design over the period that explain this; nor are they correlated with changes in, for example, changes in average cohort size or some other feature of the cohort regressions.

A more general lesson from Figure 1 is that the accuracy of DL rho estimates depends on the precise years considered. (This is also clear from DL’s results – observe the ‘relative difference (%)’ summaries reported in their Table 2 for countries with more than one year-pair of estimates – but our more complete coverage of long time periods for each country makes the finding more manifest.) For each cohort definition and head’s age range in our Table 1, there is at least one year-pair for which the DL rho estimate is very close to its ‘true’ panel counterpart. But it is rare for researchers to have access to panel data for as many year-pairs as we have, particularly in developing countries. Typically the data cover only one or two subperiods.

Our analysis shows that the quality of DL rho estimates depends on the data that happen to be available for a particular time period. Put differently, if researchers do have access to data for multiple year-pairs, then there may be a pay-off to averaging the DL rho estimates over time in order to remove what might be spurious volatility.

In the next section, we put these issues on one side, and use the estimated values of DL rho, along with the other parameters of the income models, to derive estimates of poverty dynamics statistics.

5. Synthetic panel estimates of poverty dynamics statistics: leading case

In this section we provide synthetic panel estimates of the four joint poverty probabilities and two conditional poverty probabilities, for Australia and Britain. We focus on a ‘leading case’ set of definitions, and in Section 6 consider the impact on the estimates and their validity of changes to these definitions. Our leading case is based on the combinations of sample selection criterion and cohort definition that provide estimates of DL rho that are the closest to the ‘true’ panel value (see Figure 1). This maximizes the chances that the DL methods estimates are accurate, other things being equal. Thus, the leading case is based on the following criteria: household head is aged 25–75; the cohort definition is COB*YOB(5) for Australia and Sex*YOB(5) for Britain; the poverty line is 60% of contemporary median income; and the estimates refer to ‘all individuals’.

Our leading case estimates of the four joint poverty probabilities for each year-pair are shown in Figure 2 (HILDA) and Figure 3 (BHPS), together with related benchmarks to compare them with. Figure 4 shows the estimates of poverty exit and entry probabilities for both countries. Each figure has the same format. We show DLLM parametric bounds estimates, assuming $0.5 < \rho < 0.9$. Although these ρ bounds differ from those used by DLLM for developing countries, they are consistent with the ‘true’ panel estimates that DL report for the USA (Table 2) and with Tables 1 and 2, and Figure 1 above. The black dots labelled ‘parametric estimate’ are the DL-method probability estimates, derived using the approach discussed earlier. We also show what the estimates would be were the derivation undertaken using the ‘true’ panel rho rather than DL rho. The benchmarks for assessing the accuracy of the DL-method estimates are shown by the ‘true’ estimates and their pointwise 95% confidence intervals (dark grey band).

Consider first the HILDA estimates of the joint exit probability shown in the bottom-left panel of Figure 2. The parametric bounds estimates fluctuate slightly from one year-pair to the next, but are consistently between around 4% and 9%, a range of some 5 percentage points, and hence relatively wide. The DL parametric estimates also fluctuate somewhat, but tend to lie in the middle of bounds estimates (apart from at the end of the period): the values are around 7% to 8%. If the ‘true’ panel rho had been known, the estimates of the joint probability would have been quite similar – except at the very beginning of the period and at the end of the period, which is when the DL rho and ‘true’ panel rho estimates differ the most (see Figure 1(a)). The similarities between the series are a reminder that the accuracy of the

DL-method probability estimates is also contingent on the income model predictions and related assumptions, a point to which we return in Section 7.

<Figure 2 near here>

We assess the accuracy of the DL-method estimates of the joint exit probability by considering whether they lie within the 95% confidence interval of the ‘true’ estimates. It is clear from Figure 2 that, for the vast majority of year-pairs (11 out of 14), the DL estimate is outside the 95% confidence interval of the ‘true’ estimate. On average, it is around 2 percentage points larger than the benchmark ‘true’ joint probability (more at the end of the period).

For the other three joint probabilities, the headline message regarding accuracy is mixed. For the joint persistence probability (top left hand figure in Figure 2), the DL-method estimates are accurate in the sense that they lie with the 95% confidence interval of the corresponding ‘true’ estimate in 11 out of the 14 comparisons. The DL estimates of the joint persistence probability tend to be slightly smaller than their ‘true’ counterparts, but both show a small decline from around 11% at the beginning of the 2000s to around 10% just over a decade later.

However, although the DL method estimates the joint persistence probability relatively accurately, it does not do so for the other two joint probabilities. The ‘true’ joint probability of being non-poor in two consecutive years increases by around 4 percentage points over the period, from around 76% to around 80%. By contrast, the DL estimates show no rising trend (around fluctuating values). The DL estimates are within the 95% confidence interval of the ‘true’ estimates for only four of the 14 year-pairs (all of which are at the start of the period). The joint entry probability is also inaccurately estimated, with only two of the 14 DL-method estimates within the benchmark confidence band. The ‘true’ estimate fluctuates around 5% to 6% and the DL-method estimate is somewhat larger. For 2006-2007 and the three years at the end of the period, the DL-method estimate is up to 2 percentage points greater than the upper bound of the confidence interval, which is a large gap when assessed relative to the ‘true’ point estimate.

The findings for the BHPS contain both differences and similarities to the HILDA ones: see Figure 3. On the one hand, the headline results regarding accuracy of the DL method are more favourable. For the joint persistence probability (top left hand figure), all 17 of the DL-method estimates lie within the benchmark confidence interval, and well within it in the majority of year-pairs. For each of the other three joint probabilities, around half of the DL-method estimates lie within the 95% band, and almost all of these are in the first half of

the period. As with HILDA, inaccuracy is greatest towards the end of the period and at the very beginning and, again, these are precisely the year-pairs when there are the greatest differences between DL rho and ‘true’ rho. Interestingly, the point estimates of the joint probabilities, and trends over time (or lack of trend), are similar for both Australia and Britain. The BHPS estimates tend to fluctuate a bit more over time, and the benchmark confidence interval is a little wider, with both features reflecting the smaller sample sizes in the British dataset.

<Figure 3 near here>

Our findings for the conditional probabilities are shown in Figure 4, with panel (a) displaying HILDA estimates and panel (b) the BHPS estimates. Contrary to Fields and Viollaz (2013) as cited in the Introduction, but consistent with DL, we do not find that poverty exit and entry rate estimates are markedly less accurate than joint probability estimates. Our results for conditional and joint probabilities have more similarities than differences. Benchmark confidence bands tend to be larger, especially for exit rates, reflecting the smaller sample numbers ‘at risk’ that underlie the calculations. But we also see that BHPS exit and entry rates are more accurately estimated than HILDA exit and entry rates. As well, for both countries, inaccuracies tend to be more prevalent towards the end of the time periods covered, and the DL-method estimates of exit and entry rates tend to be larger than the corresponding ‘true’ estimates.

<Figure 4 near here>

6. Synthetic panel estimates of poverty dynamics statistics: variants

Our results so far show that the accuracy of the DL method depends on the time period considered and the country context. But these findings are contingent on a number of definitional assumptions. In this section, we consider the robustness of our conclusions to variations in definitions around the leading case considered so far. For brevity, we show only a selection of our results here; the complete set is provided in the Supplementary Material.

6.1 Changing the ‘true’ panel benchmark

The DL(LM) approach is implemented at the household level, as explained in the Introduction. The time-invariant household characteristics used in the DL(LM) income

models are the characteristics of the household head, and population-level estimates (i.e. for individuals, not households) are derived using sample weights. This means that DL(LM)'s income projections are based on the implicit assumption that household composition is unchanged between Year 1 and Year 2. The characteristics of a head of household are used to predict the income of all the members of his or her household in both years. DL(LM) applications (and our Section 5) employ benchmarks in which individuals' predicted Year 1 outcomes are compared with the Year 1 income of their Year 2 household head, ignoring any changes in household head.

These benchmarks differ from what standard analysis of poverty dynamics would use when genuine longitudinal data are available. The standard approach recognises that the concept of a 'longitudinal household' cannot be defined satisfactorily because households dissolve and form over time. Only individuals can be consistently linked over time. Thus poverty dynamics according to the standard panel approach are assessed by tracking each individual in the base year sample over time and comparing their Year 1 and Year 2 household incomes. Only if there is no household change over time does the approach of tracking household heads over time lead to the same poverty dynamics estimates. But household change is prevalent and poverty entry and exit risks are correlated with household change. For a review of links between panel following rules, household change and income mobility in rich countries, see Jenkins (2011). For similar arguments in a developing country context, see Rosenzweig (2003).

In sum, taking a standard panel approach to calculating poverty dynamics statistics can potentially lead to a difference between DL's 'true' panel estimates and 'standard' panel estimates of poverty statistics which are based on data derived by following individuals, not household heads. But does the change of benchmark affect assessments of the validity of the synthetic panel approach?

We answer this question focusing on estimates of poverty exit and entry rates. Figure 5 displays estimates of these for HILDA (panel a) and BHPS (panel b). The figures have exactly the format as Figure 4, except that we have removed the parametric bounds for clarity, and added the alternative 'standard' panel estimates and their pointwise 95% confidence bands. These are shown in light grey and may be contrasted with the DL 'true' panel benchmark shown in dark grey. We also include the estimates derived applying the standard panel approach to data on all individuals in all households regardless of head's age that thus cover the whole population (long-dashed line labelled 'all hh').

<Figure 5 near here>

Figure 5 shows that the standard panel approach leads to estimates of poverty exit rates that are slightly larger in magnitude and more noticeably larger estimates of poverty entry rates. Using the standard panel benchmarks leads to a more favourable assessment of the accuracy of DL-method synthetic panel estimates of poverty entry rates. For HILDA, the number of year-pair estimates lying outside the reference confidence band falls from 12 out of 14 ('true' panel benchmarks) to 5 ('standard' panel benchmarks). For the BHPS estimates of poverty entry rates, the corresponding fall is from 9 out of 17 to 4.

Whether changing the benchmark definition would make a difference in other contexts is difficult to assess. Taking account of household change is likely to raise poverty entry rate estimates rather than exit rate estimates in most countries compared to the 'true' panel approach. This is consistent with the well-known finding for rich countries that household change is particularly associated with poverty entries rather than poverty exits (Bane and Ellwood 1986, Jenkins 2011). However, synthetic panel estimates will only be assessed as more accurate according to the standard panel benchmarks if, as well, the synthetic estimates are larger than the 'true' benchmark. Although this is the situation in our datasets, it may not always be the case, as shown by the case of the USA in DL's Table 5 (the synthetic panel entry rate estimate is 4.4%; the 'true' estimate is 5%).

We leave open for further research the issue of what is the appropriate benchmark to use to assess the accuracy of synthetic panel estimates and, for the rest of this paper, we return to using the 'true' panel estimates given our DL reference point.

6.2 Changing the household head's age range

We now consider the impact of narrowing the age range of the household head from 25–75 to 25–55, retaining all other definitions associated with our leading case. Figure 6 displays estimates for HILDA (panel a) and BHPS (panel b) and is directly comparable with Figure 4 (based on the wider household head age range).

<Figure 6 near here>

Using the narrower age range is associated with a substantial deterioration in the accuracy of the HILDA estimates. Now none of the DL-method point estimates of the poverty exit rate lie within the reference 95% confidence band, and only one of the entry rate estimates does. Indeed several of the point estimates lie outside the corresponding upper bound estimate. The poorer quality of the estimates can be traced back to the poorer accuracy estimate of the DL rho estimates in this case: compare the 'Parametric est.' and 'Parametric

est. true rho' series in Figure 6 and see also the two corresponding bottom figures in Figure 1(a).

For the BHPS, use of the narrower age range also leads to poorer quality estimates of poverty exit and entry rates, but the effect is not nearly as marked as in the Australian case. The deterioration in quality is also related to the poorer accuracy of DL rho estimates relative to the leading case, but the effects are not as large as for HILDA (Figure 1, panel (b)).

The Supplementary Material contains a complete collection of figures showing the impact of using the 25–55 age range compared to the 25–75 one, and for joint probabilities as well as the conditional probabilities discussed here. The appendices confirm that changing the definitions away from the leading case scenario generally leads to less accurate estimates of all poverty dynamics statistics.

6.3 Changing the poverty line to 50% of the contemporary median

We now consider the impact of changing the poverty line to 50% of contemporary national median income (from 60%), but otherwise retaining all other leading case definitions. Figure 7 shows the joint probability estimates for HILDA and Figure 8 shows them for the BHPS. Figure 9 shows the estimates of poverty exit and entry probabilities for both countries. Compare these figures with Figures 2, 3 and 4, respectively, for the leading case.

The most important finding is that, with the lower poverty line, the DL method estimates for Australia are more accurate. For example, the number of estimates of the joint exit probability that lie within benchmark 95% confidence interval increases from 3 out of 14 to 10. For the joint persistently non-poor probability, the corresponding numbers are an increase from 3 to 9 and, for the joint entry probability, an increase from 2 to 7. There are corresponding improvements in the accuracy of the DL-method estimates of poverty exit rates (the number within the reference confidence band increasing from 2 to 10) and of poverty entry rates (the number increasing from 2 to 7): see Figure 9(a).

<Figures 7, 8, and 9 near here>

There is also some increase in estimate accuracy associated with using the lower poverty line in the British case though the effect is less noticeable. Indeed, the joint poverty persistence probability is now less accurately estimated by the DL method, with number of estimates lying in the reference 95% confidence band falling to 9/17 compared to 17/17 in the leading case scenario. There is no change in the number of estimates of the joint non-poverty persistence probability within the reference band (9/17), but the number increases to 11 from

9 for the joint exit probability and to 13 from 8 for the joint entry probability. See Figure 8. The accuracy of the DL-method estimates of poverty exit rates also improves compared to the leading case scenario (the number within the reference confidence band increases from 9 to 14) and of poverty entry rates (the number increases from 9 to 13): see Figure 9(b).

Our analysis demonstrates that the accuracy of the DL-method estimates of poverty dynamics statistics is sensitive to the choice of poverty line. DLLM undertook extensive analysis of the robustness of their non-parametric bounds method to a wide range of poverty lines, for Indonesia and also Vietnam, and they conclude that ‘our approach is found to work well for the full possible range of poverty lines that might be specified’ (DL: 122). Cruces et al. (2015, section 5.4) report similar findings for Peru, Chile, and Nicaragua. By working well, the authors mean that ‘true’ panel estimate lies between the upper and lower bound estimates regardless of the poverty line chosen. We find this result too, as our Figures show. But our new finding concerns the sensitivity of the DL method’s point estimates. It appears that going beyond estimation of bounds to derive point estimates also runs the risk of lack of robustness. We have focused on poverty line specifications that are in common use in rich countries. An interesting task for future research is analysis of the robustness of the DL method to poverty line choice in developing country settings.

6.4 Estimates for population subgroups

DL argue that their method provides good estimates for population subgroups as well as the population as a whole, using regional breakdowns to illustrate their case. Here we consider the accuracy of subgroup estimates using breakdowns by age, reflecting rich country policy interest. Figure 10 shows estimates of poverty exit rates and entry rates for HILDA (panel a) and the BHPS (panel b) for individuals aged 18–59 rather than all individuals in households with heads aged 25–75 (as in Figure 4). For corresponding estimates for individuals aged 0–17, and aged 60–75, see the Supplementary Material, which also shows subgroup estimates of joint probabilities.

The HILDA estimates of poverty exit rates for the 18–59 subgroup are noticeably more accurate than those for all individuals, with 10/14 estimates within the 95% confidence interval around the ‘true’ estimates. The figures in the Supplementary Material show that the poor performance of the estimates for all individuals (only 3/14 estimates within the reference band; see Figure 4) is due to the poor accuracy of the estimates for the 60–75 subgroup (and not the 0–17 subgroup). The HILDA estimates of the poverty entry rates for

the 18–59 subgroup are less accurate than those for all individuals. The number of estimates within the benchmark band is 1/14 (rather than 2/14) and the synthetic panel estimates overestimate the ‘true’ panel estimates by a greater amount. In contrast, the estimates for the 0–17 and 60–75 subgroups are noticeably more accurate than the estimates for all individuals.

<Figure 10 near here>

The BHPS results have similarities to the HILDA ones. The estimates of poverty exit rates for individuals aged 18–59 are slightly more accurate than the corresponding estimates for all individuals (14/17 estimates within the benchmark confidence interval rather than 10/17) but, by contrast with the HILDA estimates, the inaccuracy in the estimates for all individuals is accounted for by the estimates for the 0–17 subgroup rather than the 60–75 subgroup. The BHPS estimates of the poverty entry rates for the 18–59 subgroup are less accurate than their counterparts for all individuals, with the number of estimates within the benchmark confidence interval being only 1/17 (compared with 8/17). As for HILDA, the estimates for the 0–17 and 60–75 subgroups are more accurate than the estimates for all individuals.

Our overall conclusion regarding the accuracy of subgroup estimates is more equivocal than DL’s. In our analysis, some of the estimates for some subgroups and for some statistics (conditional or joint probabilities) are more accurate than the corresponding estimates for all individuals. At the same time, some are noticeably less accurate.

7. Conclusions

Our analysis shows that the DL method performs less well when applied to Australian and British data than it does in previous studies using data for middle- and low-income countries. To what extent are our findings about validity generally applicable and to what extent do they arise from having analysed two particular rich countries for which high-quality panel data exist?

Clearly, our findings about the potentially poor accuracy of synthetic panel estimates are less of an issue in contexts where there are no panel data at all. As DL point out, ‘... this basic methodology offers significant potential towards a better understanding of poverty dynamics in *settings where panel data are absent* and can serve as *a rather promising avenue for further research*’ (p. 37; emphasis added). Thus, synthetic panel estimates have potential

in many developing countries, and also in rich country contexts where panel data are unavailable or of doubtful quality.

Our research shows that there is scope for further research that is relevant for developing as well as developed countries. For example, it would be useful to know whether some of the sensitivities in the DL method sensitivities we have found, such as to the choice of head's age range and to the poverty line, are also present in other country contexts. We also need further research comparing the DL method with other synthetic panel approaches such as that proposed by Bourguignon and Moreno (2015). Garcés Urzainqui (2017) is the only study to do this to date.

Aside from data availability and quality issues, there is the question of whether the synthetic panel approach is less applicable to high-income countries rather than middle- or low-income countries because the underlying assumptions are less appropriate or the income modelling does not work so well in high-income country contexts (or, related, the nature of the income distribution around the poverty line is different in high-income countries). In this connection we note that, of the five countries used in DL's validation study, the synthetic panel estimates are slightly less accurate for the one rich country they included (USA): compare the 'goodness of fit' statistics in their Tables 3–5.

Key assumptions in DLLM's parametric bounds and DL's point estimate approaches are that Year 1 and Year 2 incomes are bivariate lognormally distributed and that income predictors are time-invariant (see Section 2). We have checked the lognormality of the marginal distributions and it appears that this issue is no greater a problem in HILDA and the BHPS than it was for DL's study countries (see the Supplementary Material, Appendices A3, B3). We have also compared bivariate densities of model residuals with bivariate densities of standard normal distributions with the same ρ as the estimated one. Contour plots for each country year-pair are quite similar (see the Supplementary Material, Appendices A5, B5). Nevertheless, it may be that a different approach from bivariate normality, e.g. involving copula specifications for dependence, is more fruitful.

Regarding the success of the income modelling in different country contexts, we note that the adjusted- R^2 of our income regressions based on HILDA and BHPS data are lower than the adjusted- R^2 that DLLM and DL report. (Cruces et al. 2015 do not report R^2 .) For example, DL report adjusted- R^2 for the USA between 0.29 and 0.34 (Appendix 2, Table 2.1) whereas our adjusted- R^2 for Australia and Britain are around two-thirds of the US values (see the Supplementary Material), although the income predictors are much the same in all three

countries.

The income predictions are likely to be one source of the different validity findings, but they are unlikely to be the main one. DL undertake simulation analysis comparing estimate validity for a scenario in which analysts have access to the full set of income predictor variables with other scenarios in which analysts have fewer variables (the bivariate normality assumption is maintained). DL (p. 27) report that, even when few time-invariant variables are available, the synthetic panel estimates compare favourably with their true counterparts (as long as the sample size is not very large – larger than the samples we have used in our study). DL’s simulation finding is consistent with our finding (cited in Section 3) that using a smaller set of income predictors led to very similar synthetic panel estimates and hence assessments of accuracy. More extensive simulation analysis, building in a number of other departures from the DL modelling assumptions (including bivariate lognormality), may help unravel which factors contribute most to the differences in synthetic panel estimate validity across countries.

Although our study focuses on the DL method, we have also derived DLLM parametric bounds estimates for the same sets of assumptions and definitions. The important finding is that the bounds virtually always include the ‘true’ panel estimate for each poverty dynamics statistic, year, and country. (The exceptions concern the joint non-poor persistence probability: see Figures 2 and 3. But even in these cases the 95% confidence interval includes the upper bound.) This suggests, first, that researchers employing the DL (or related) method should always show DLLM parametric bounds estimates as well as point estimates. Second, there may be significant returns to investments in getting estimates of ρ from external sources that are as credible and as narrow as possible.

Acknowledgements

We dedicate this paper to the memory of Tony Atkinson. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne

Institute. Our research is supported by an Australian Research Council Discovery Grant (award DP150102409). Jenkins's research is also partially supported by core funding of the Research Centre on Micro-Social Change at the Institute for Social and Economic Research by the University of Essex and the UK Economic and Social Research Council (award ES/L009153/1). For helpful discussions, we thank Hai-Anh Dang, Peter Lanjouw, David Garcés Urzainqui, the handling editor (Markus Jäntti), and an anonymous referee. We thank DLLM for making their Stata code freely downloadable. Our Stata code, which builds on theirs, is available on request. Helpful comments from audiences in Bristol, Dublin, Essex, Melbourne, and Oslo are also acknowledged.

References

- Bane, M. and Ellwood, D. Slipping into and out of poverty: the dynamics of spells. *Journal of Human Resources*, 21: 1–23, 1986.
- Bourguignon, F. and Moreno, H. On the construction of synthetic panels. Paper presented at the North East Universities Development Consortium annual conference, Brown University, Providence RI, 2015.
- Canberra Group. Handbook on Household Income Statistics, second edition. Geneva: United Nations Economic Commission for Europe, 2011.
- Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., and Viollaz, M. Estimating poverty transitions using repeated cross-sections: a three-country validation exercise. *Journal of Economic Inequality*, 13: 161–179, 2015.
- Dang, H.-A. and Dabalen, A. L. Is poverty in Africa mostly chronic or transient? Evidence from synthetic panel data. *Journal of Development Studies*, online, 2018. DOI: 10.1080/00220388.2017.1417585
- Dang, H.-A. and Ianchovichina, E. Welfare dynamics with synthetic panels: the case of the Arab world in transition. *Review of Income and Wealth*, 64 (S1): S114–S144, 2018.
- Dang, H.-A. and Lanjouw, P. Measuring poverty dynamics with synthetic panels based on cross-sections. Policy Research Working Paper 6504, The World Bank, 2013.
- Dang, H.-A. and Lanjouw, P. Welfare dynamics measurement: two definitions of a vulnerability line and their empirical application. *Review of Income and Wealth*, 63(4): 633–660, 2017.

- Dang, H.-A. and Lanjouw, P. Poverty dynamics in India between 2004 and 2012: Insights from longitudinal analysis using synthetic panel data. *Economic Development and Cultural Change*, 67(1): 131–170, 2018.
- Dang, H.-A., Lanjouw, P., Luoto, L., and McKenzie, D. Using repeated cross-sections to explore movements into and out of poverty. *Journal of Development Economics*, 107: 112–128, 2014.
- Dang, H.-A., Lanjouw, P., and Swinkels, R. Who remained in poverty, who moved up, and who fell down? An investigation of poverty dynamics in Senegal in the 2000s. In: M. Nissanke and M. Ndulo (eds), *Poverty Reduction in the Course of African Development*. Oxford: Oxford University Press, 2017.
- Deaton, A. Panel data from time series of cross-sections. *Journal of Econometrics*, 30: 109–126, 1985.
- Ferreira, F. H. G., Messina, J., Rigolini, J., López-Calva, L.-F., Lugo, M. A., and Vakis, R. *Economic Mobility and the Rise of the Latin American Middle Class*. Washington DC: The World Bank, 2013.
- Fields, G. and Viollaz, M. Can the limitations of panel datasets be overcome by using pseudo-panels to estimate income mobility? Paper presented at the ECINEQ Conference, Bari, Italy, 2013.
- Frick, J. R., Jenkins, S. P., Lillard, D. R., Lipps, O., and Wooden, M. The Cross-National Equivalent File (CNEF) and its member country household panel studies. *Schmollers Jahrbuch – Journal of Applied Social Sciences Studies*, 127 (4): 627–654, 2007.
- Garcés Urzainqui, D. Poverty transitions without panel data? An appraisal of synthetic panel methods. Paper presented at the ECINEQ Conference, New York City, 2017.
- Jenkins, S. P. *Changing Fortunes. Income Mobility and Poverty Dynamics in Britain*. Oxford: Oxford University Press, 2011.
- Jenkins, S. P. and Van Kerm, P. How does attrition affect estimates of persistent poverty rates? The case of EU-SILC. In: A. B. Atkinson, A. Guio, and E. Marlier (eds), *Monitoring Social Europe, 2017 Edition*. Luxembourg: Eurostat, 401–417, 2017.
- Moffitt, R. Identification and estimation of dynamic models with a time series of repeated cross sections. *Journal of Econometrics*, 59: 99–123, 1993.
- OECD. *Employment Outlook 2015*. Paris: OECD Publishing, 2015.
- OECD. *A Broken Elevator? How to Promote Social Mobility*. Paris: OECD Publishing, 2018.

- Perez, V. Moving in and out of poverty in Mexico: what can we learn from pseudo-panel methods? ISER Working Paper 2015-16, University of Essex, 2015.
- Rama, M., Béteille, T., Li, Y., Mitra, P. K., and Newman, J. L. Addressing Inequality in South Asia. Washington DC: The World Bank, 2014.
- Rigolini, J., Vakis, R., and Lucchetti, L. Left Behind: Chronic Poverty in Latin America and the Caribbean. Washington DC: The World Bank, 2016.
- Rosenzweig, M. R. Payoffs from panels in low-income countries: economic development and economic mobility, *American Economic Review Papers and Proceedings*, 93(2): 112–117, 2003.
- Summerfield, M., Freidin, S., Hahn, M., La, N., Li, N., Macalalad, N., O’Shea, M., Watson, N., Wilkins, R., and Wooden, M. HILDA User Manual – Release 15. Melbourne: Melbourne Institute for Applied Social and Economic Research, 2016.
- Verbeek, M. Synthetic panels and repeated cross-sections. In: L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data*. Berlin: Springer-Verlag, 2008, 369–383.
- Watson, N. and Wooden, M. Re-engaging with survey non-respondents: the BHPS, SOEP and HILDA survey experience, HILDA Project Discussion Paper 1/11, 2011.
- Wilkins, R. The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 15. Melbourne: Melbourne Institute of Applied Social and Economic Research, 2017.

Table 1. Estimates of ρ (year-pair averages), by cohort definition and household head's age range: (a) HILDA

	Cohort definition						
	YOB(1)	YOB(5)	Sex* YOB(5)	Sex* YOB(10)	COB* YOB(3)	COB* YOB(5)	COB* YOB(10)
<i>(i) Sample: household head aged 25–55</i>							
Panel rho	0.74 (0.03)	0.74 (0.03)	0.74 (0.03)	0.74 (0.03)	0.74 (0.03)	0.74 (0.03)	0.74 (0.03)
DL rho	0.14 (0.19)	0.61 (0.29)	0.55 (0.20)	0.74 (0.19)	0.35 (0.21)	0.49 (0.21)	0.67 (0.19)
No. of cohorts	31	6	12	6	20	16	9
Av. cohort size	135	689	345	686	318	467	901
Cohort R^2	0.02	0.01	0.02	0.01	0.02	0.02	0.02
<i>(ii) Sample: household head aged 25–75</i>							
Panel rho	0.73 (0.02)	0.73 (0.02)	0.73 (0.02)	0.73 (0.02)	0.73 (0.02)	0.73 (0.02)	0.73 (0.02)
DL rho	0.54 (0.10)	0.87 (0.07)	0.78 (0.08)	0.88 (0.07)	0.64 (0.11)	0.71 (0.09)	0.82 (0.08)
No. of cohorts	51	10	20	10	32	27	15
Av. cohort size	123	621	312	619	304	404	781
Cohort R^2	0.05	0.04	0.05	0.04	0.05	0.05	0.05

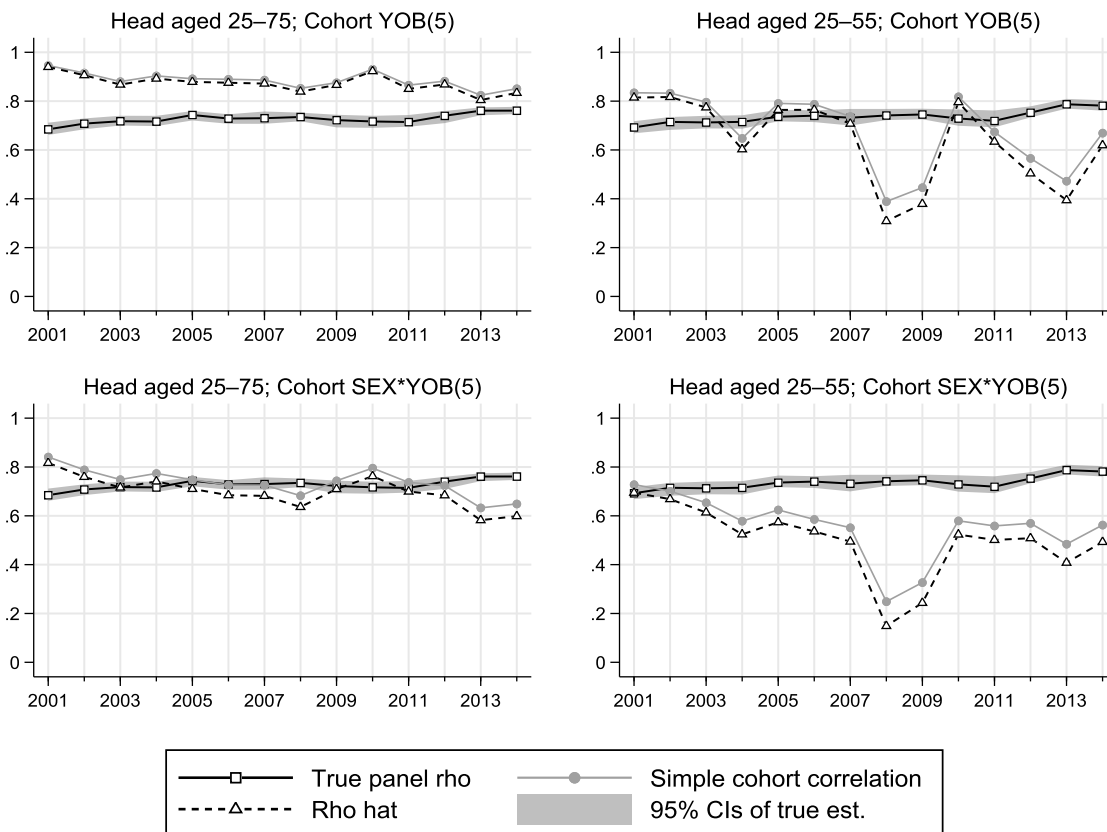
Notes. The rho estimates are the average of the estimates for each of the 14 consecutive year-pairs of HILDA data (2001/2002 to 2014/2015). The estimates for each year-pair are based on disjoint subsamples representing averages over 50 sample splits. Shown in parentheses are the standard deviations of the corresponding rho estimates, calculated over all 50 sample splits and 14 year-pairs. Each column represents a different cohort definition, where YOB(s) = year of birth (at intervals of s years), COB = Country of birth (Australia, English speaking country, non-English-speaking country). All estimates are weighted and adjusted for survey design, with the exception of the number of cohorts and average cohort size (which are unweighted). Panel rho: 'true' correlation between residuals of Year 1 and Year 2 income model (see eqn. 3 in main text). Cohort R^2 : adjusted R^2 of the regression of income on cohort variables.

Table 2. Estimates of ρ (year-pair averages), by cohort definition and household head's age range: (b) BHPS

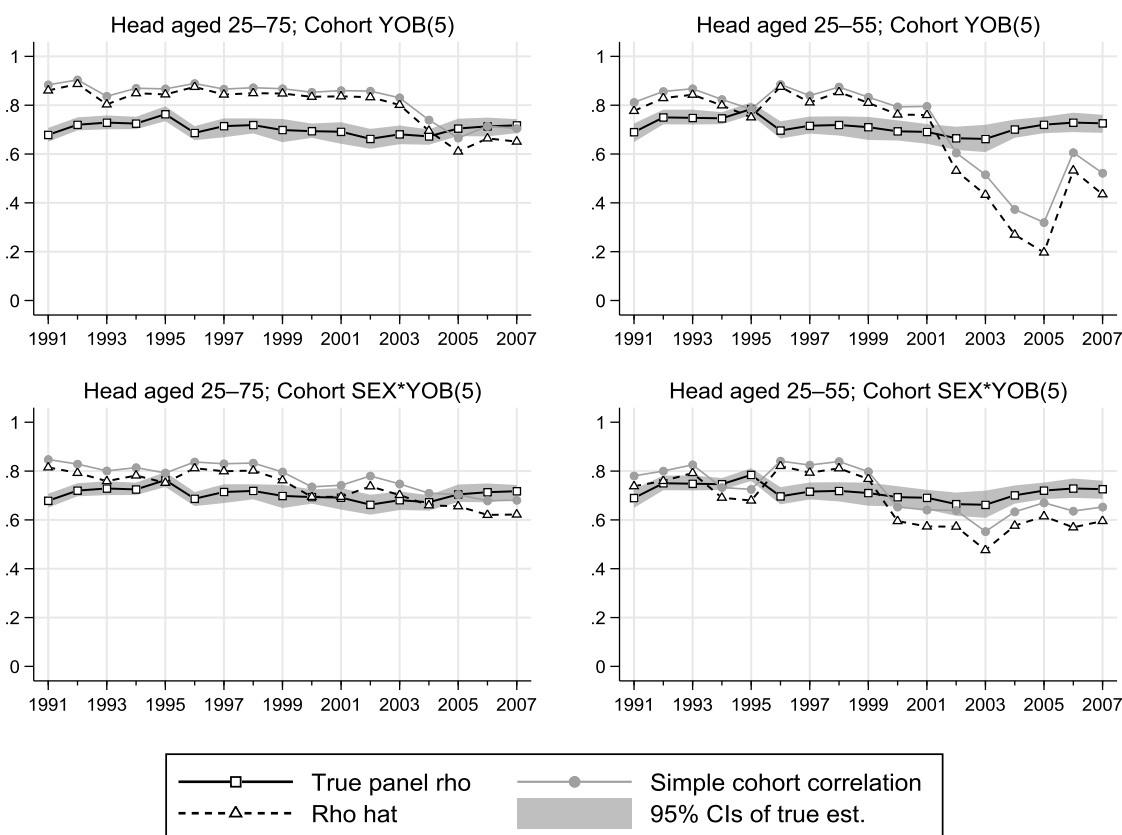
	Cohort definition					
	YOB(1)	YOB(3)	YOB(5)	YOB(10)	Sex*YOB(5)	Sex*YOB(10)
<i>(i) Sample: household head aged 25–55</i>						
Panel rho	0.71 (0.04)	0.71 (0.04)	0.71 (0.04)	0.71 (0.04)	0.71 (0.04)	0.71 (0.04)
DL rho	0.23 (0.24)	0.52 (0.30)	0.65 (0.30)	0.87 (0.20)	0.68 (0.16)	0.84 (0.13)
No. of cohorts	28	10	6	3	12	6
Av. cohort size	72	194	323	641	174	345
Cohort R^2	0.03	0.02	0.02	0.02	0.04	0.04
<i>(ii) Sample: household head aged 25–75</i>						
Panel rho	0.71 (0.03)	0.71 (0.03)	0.71 (0.03)	0.71 (0.03)	0.71 (0.03)	0.71 (0.03)
DL rho	0.43 (0.17)	0.69 (0.16)	0.79 (0.13)	0.89 (0.10)	0.73 (0.11)	0.85 (0.08)
No. of cohorts	42	17	10	5	20	10
Av. cohort size	75	181	297	590	159	315
Cohort R^2	0.05	0.04	0.04	0.03	0.06	0.05

Notes. The rho estimates are the average of the estimates for each of the 17 consecutive year-pairs of BHPS data (1991/1992 to 2007/2008). The estimates for each year-pair are based on disjoint subsamples representing averages over 50 sample splits. Shown in parentheses are the standard deviations of the corresponding rho estimates, calculated over all 50 sample splits and 17 year-pairs. Each column represents a different cohort definition, where YOB(s) = year of birth (at intervals of s years). All estimates are weighted and adjusted for survey design, with the exception of the number of cohorts and average cohort size (which are unweighted). Panel rho: 'true' correlation between residuals of Year 1 and Year 2 income model (see eqn. 3 in main text). Cohort R^2 : adjusted R^2 of the regression of income on cohort variables.

Figure 1. Year-by-year variation in estimates of ρ , by cohort definition and household head's age range
(a) HILDA

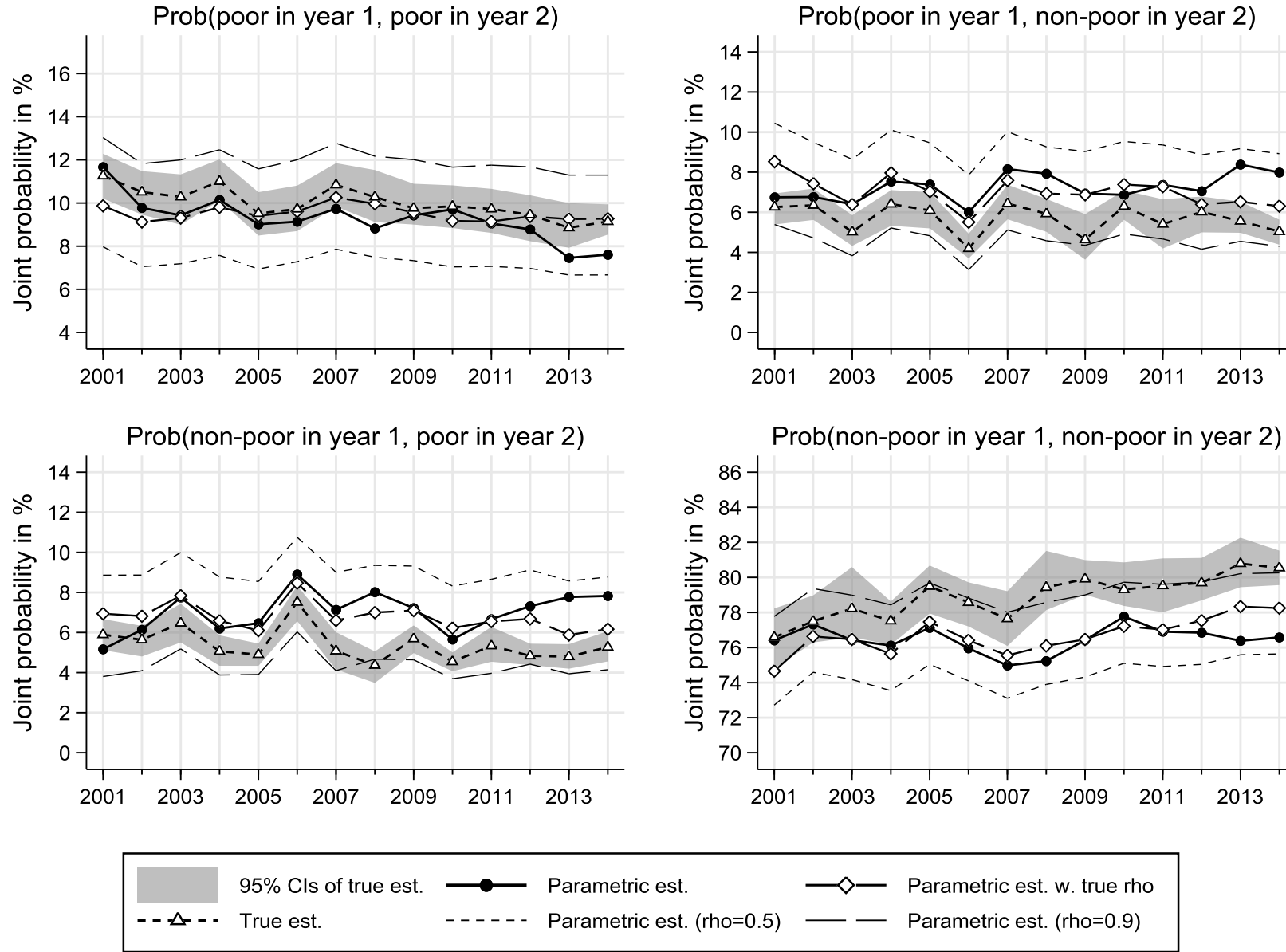


(b) BHPS



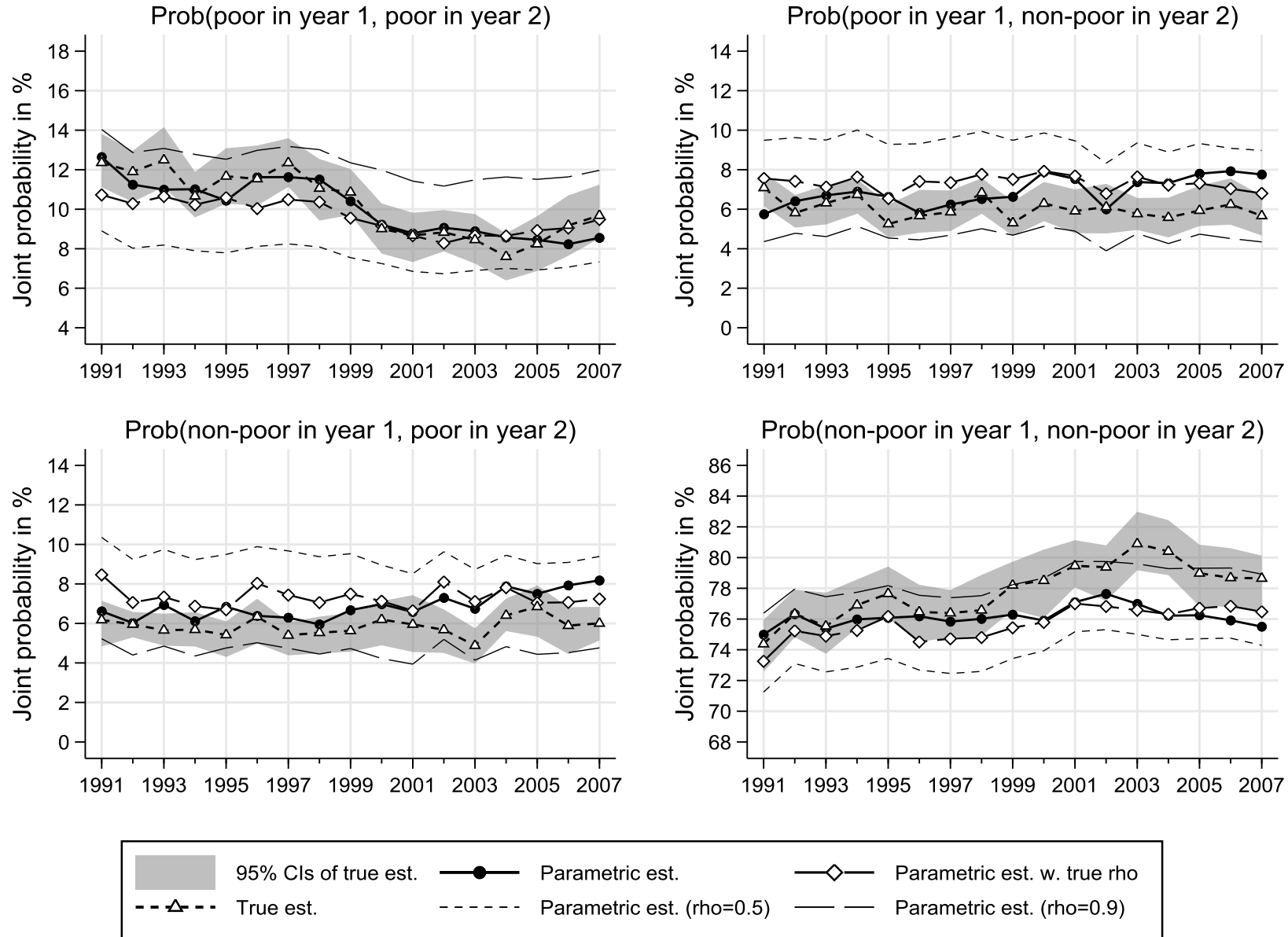
Notes. See main text, and Tables 1 and 2, for explanations of the cohort definitions.

Figure 2. Estimates of joint poverty probabilities, by year ('leading case' definitions): HILDA



Notes. Leading case definitions: household head aged 25–75 years in Year 1, cohort definition COB*YOB(5), poverty line is 60% of contemporary median income. Estimates refer to all individuals living in households with a head aged 25–75.

Figure 3. Estimates of joint poverty probabilities, by year ('leading case' definitions): BHPS



Notes. Leading case definitions: household head aged 25–75 years in Year 1, cohort definition SEX*YOB(5), poverty line is 60% of contemporary median income. Estimates refer to all individuals living in households with a head aged 25–75.

Figure 4. Estimates of poverty exit and entry probabilities, by year ('leading case' definitions)

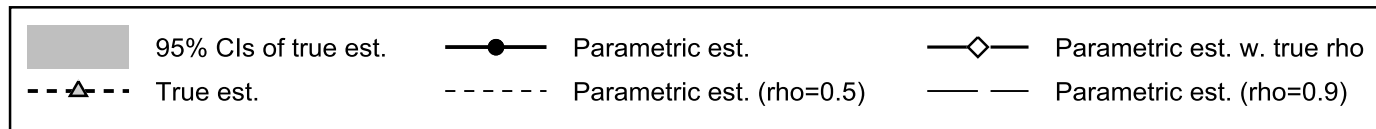
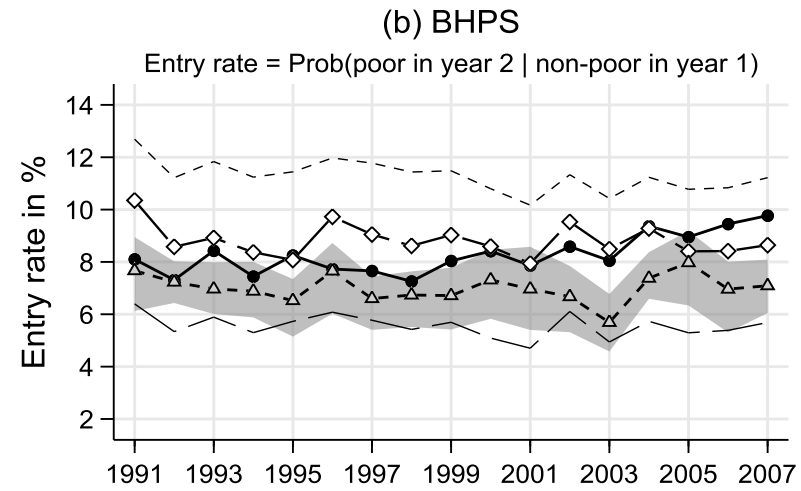
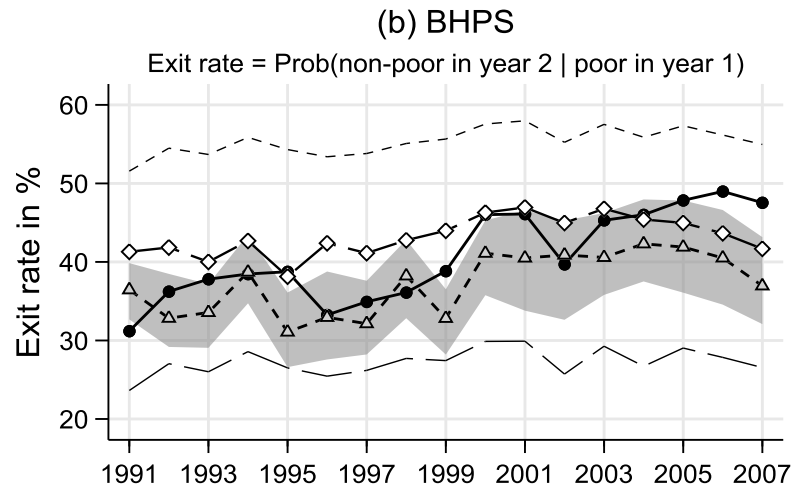
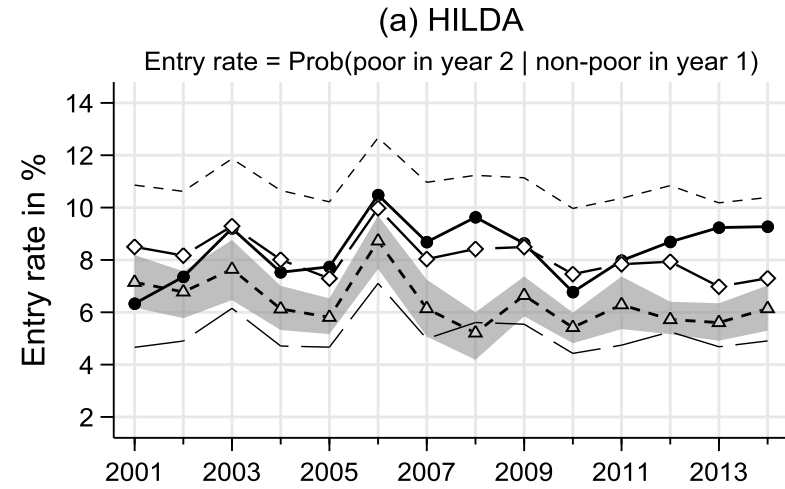
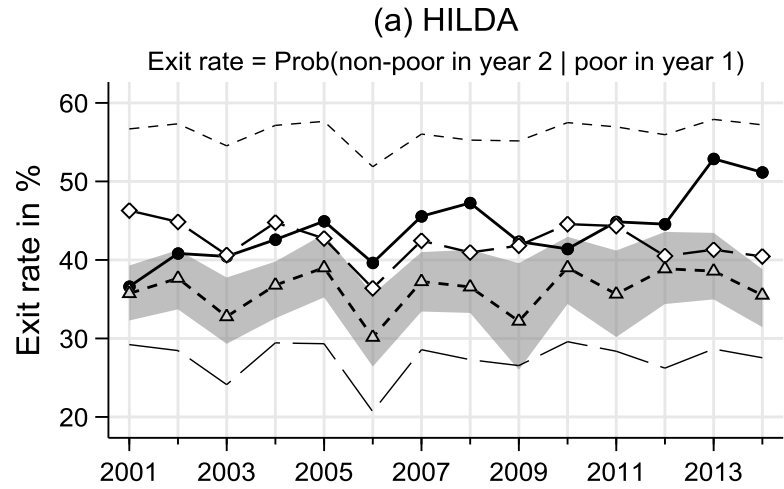


Figure 5. 'True' versus 'standard panel' estimates of poverty exit and entry probabilities, by year ('leading case' definitions)

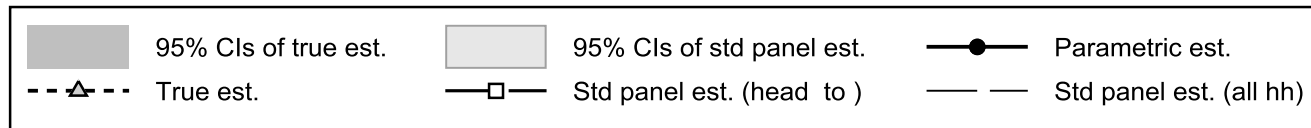
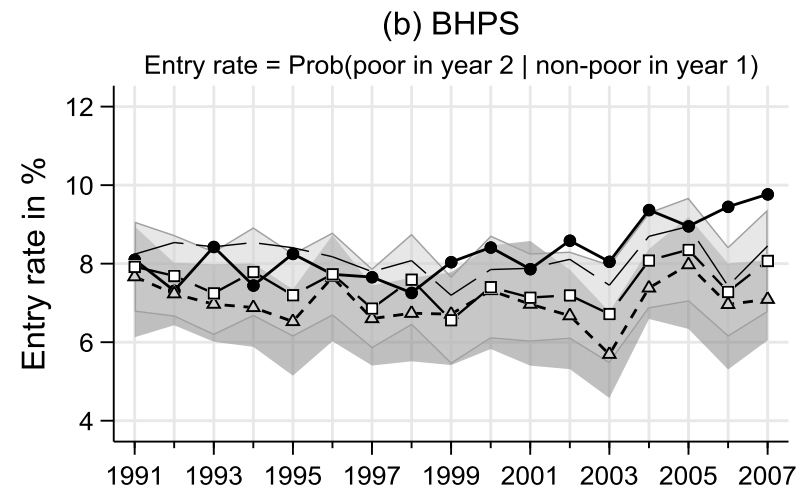
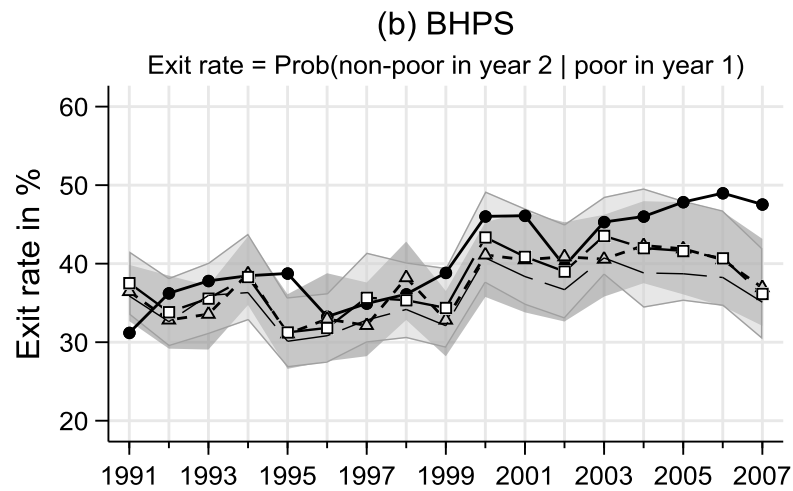
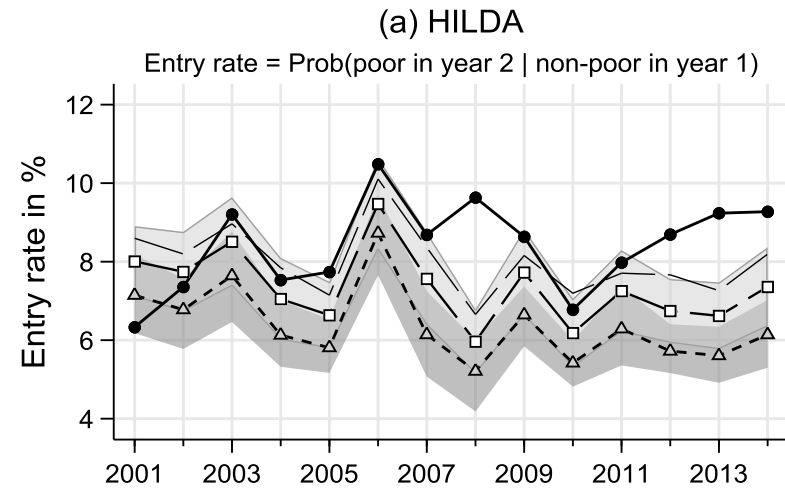
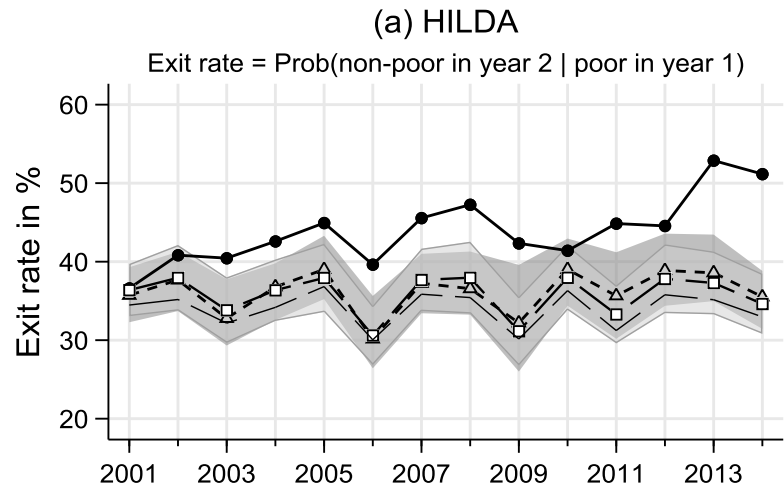


Figure 6. Estimates of poverty exit and entry probabilities, by year, household head aged 25–55 ('leading case' definitions otherwise)

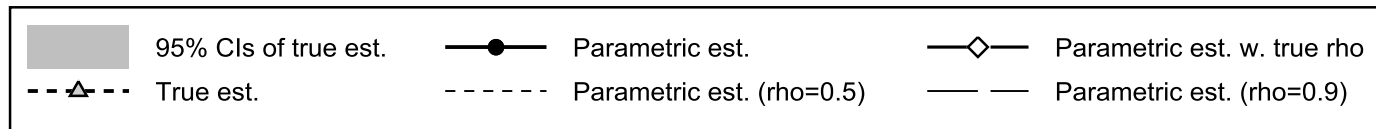
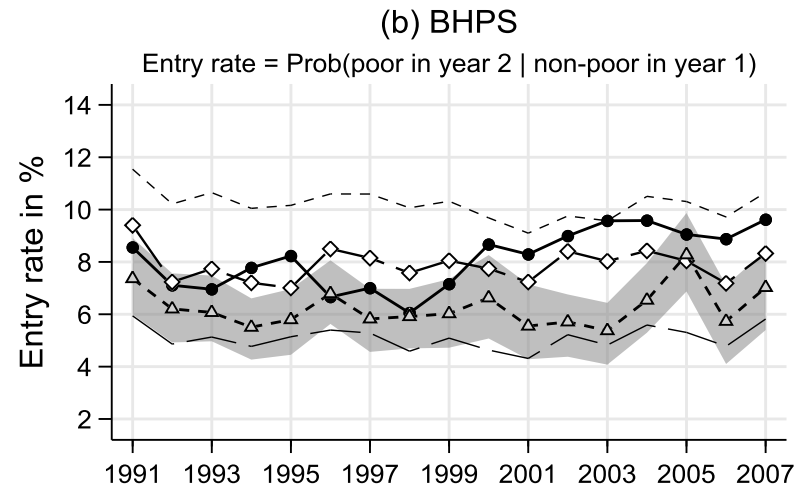
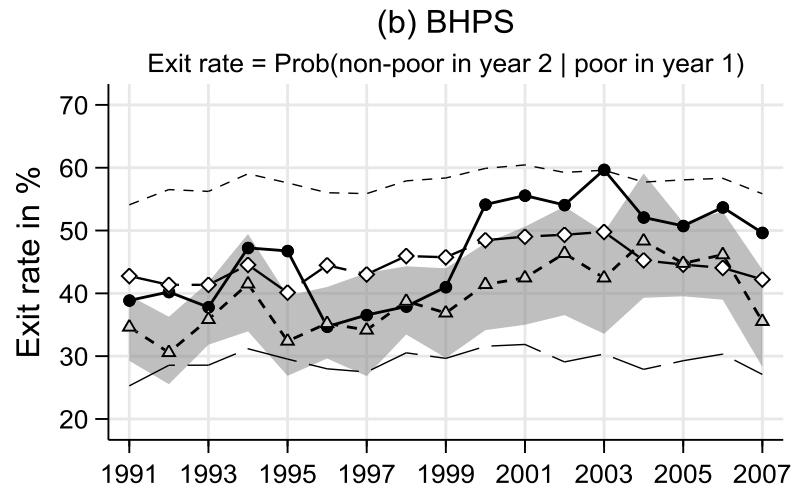
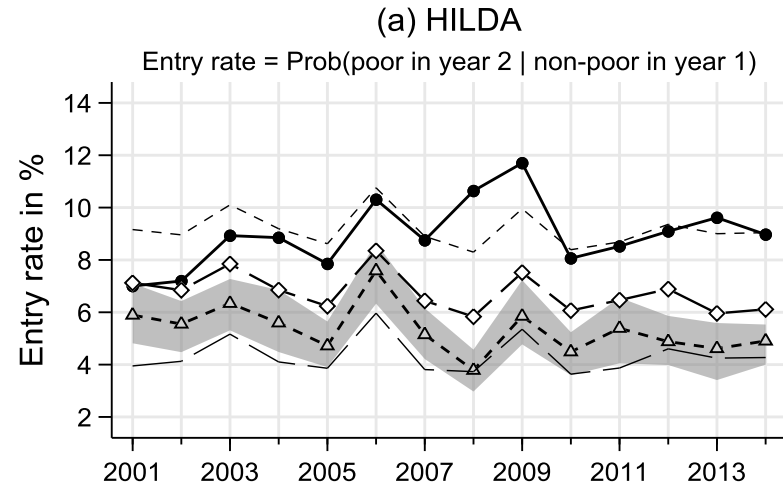
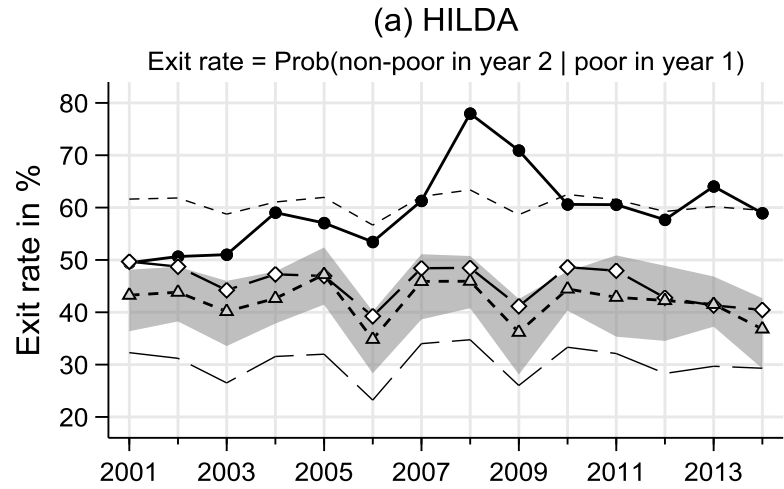


Figure 7. Estimates of joint poverty probabilities, by year, poverty line is 50% contemporary median ('leading case' definitions otherwise): HILDA

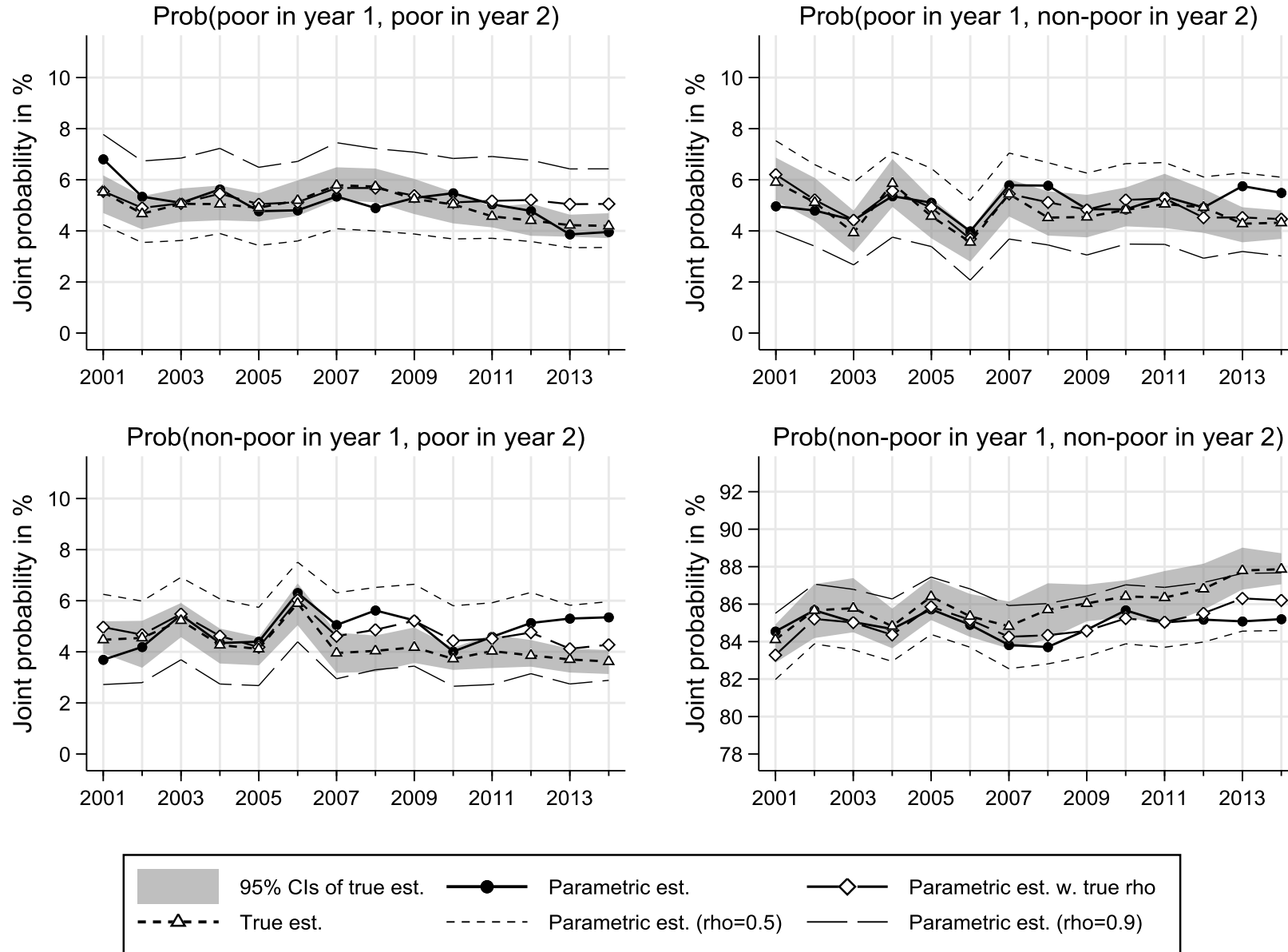


Figure 8. Estimates of joint poverty probabilities, by year, poverty line is 50% contemporary median ('leading case' definitions otherwise): BHPS

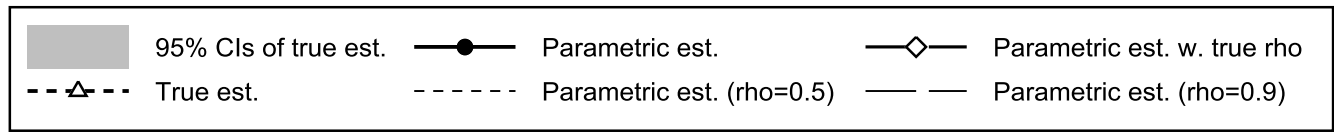
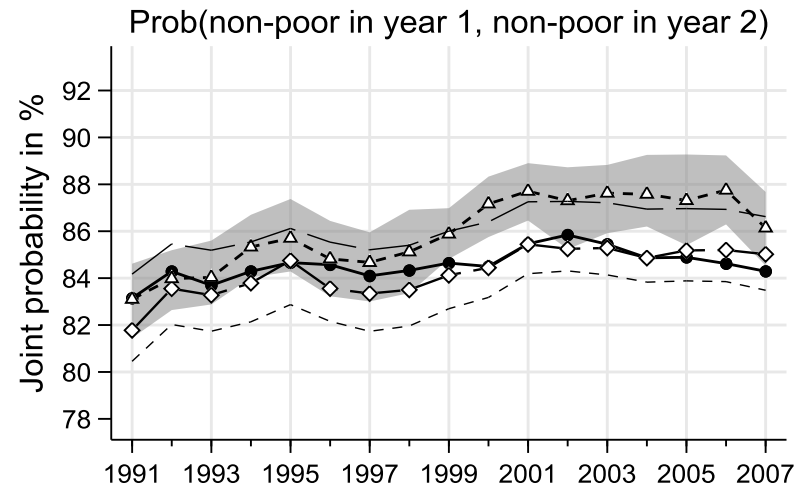
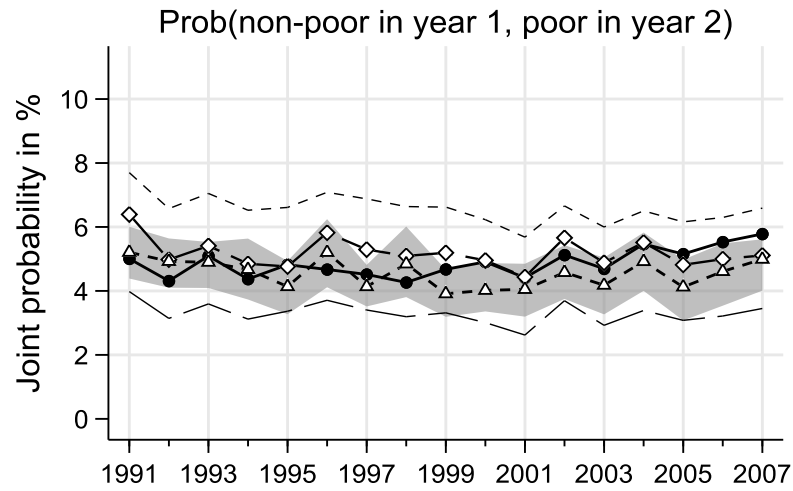
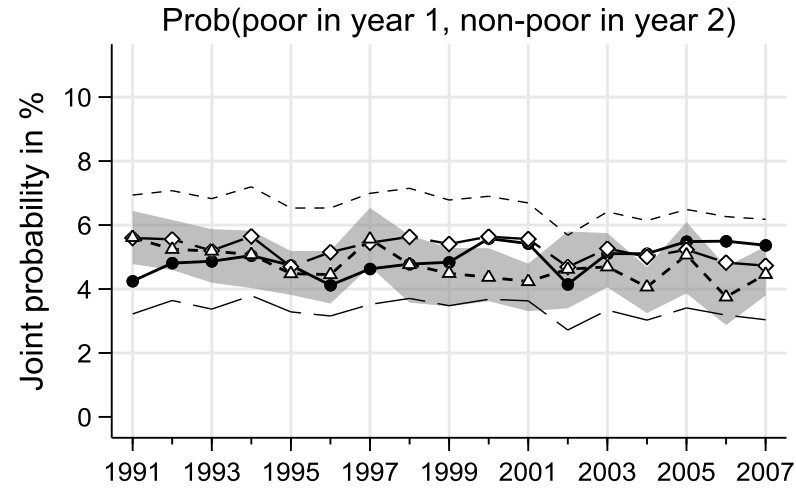
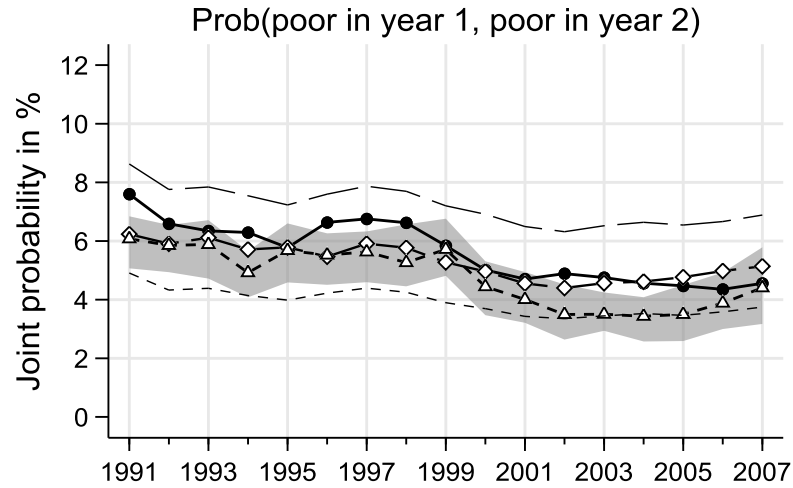


Figure 9. Estimates of poverty exit and entry rates, by year, poverty line is 50% contemporary median ('leading case' definitions otherwise)

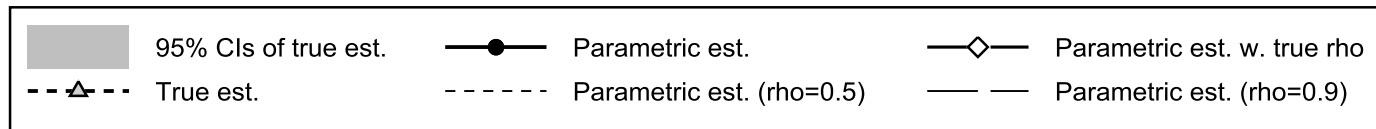
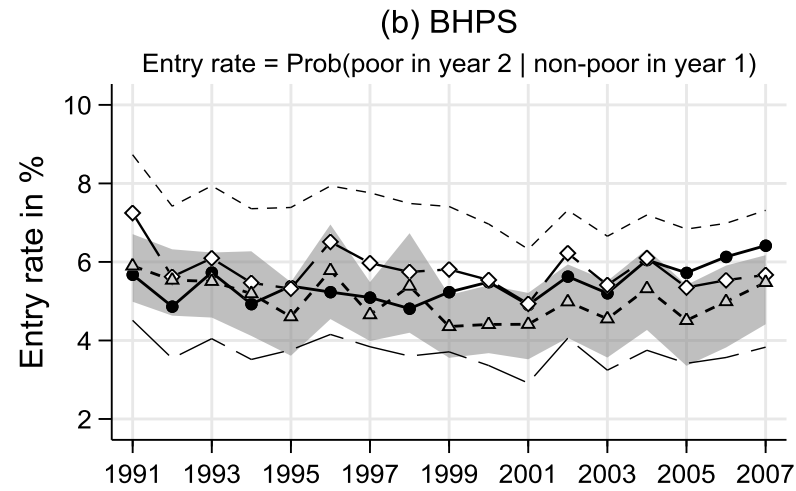
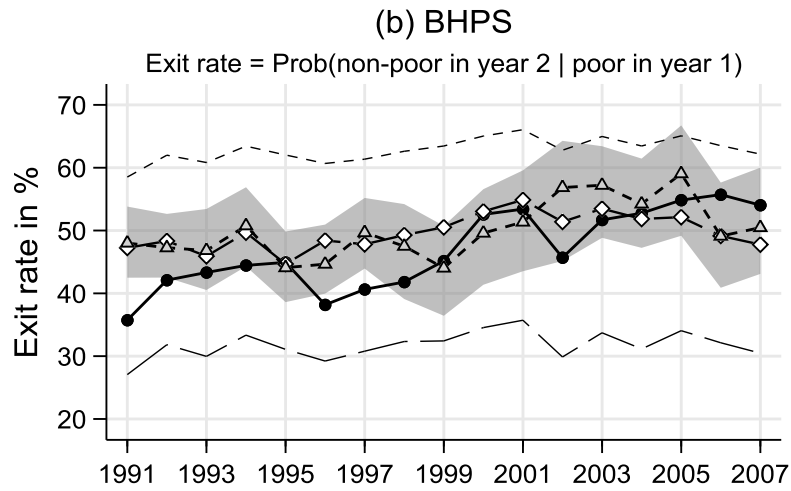
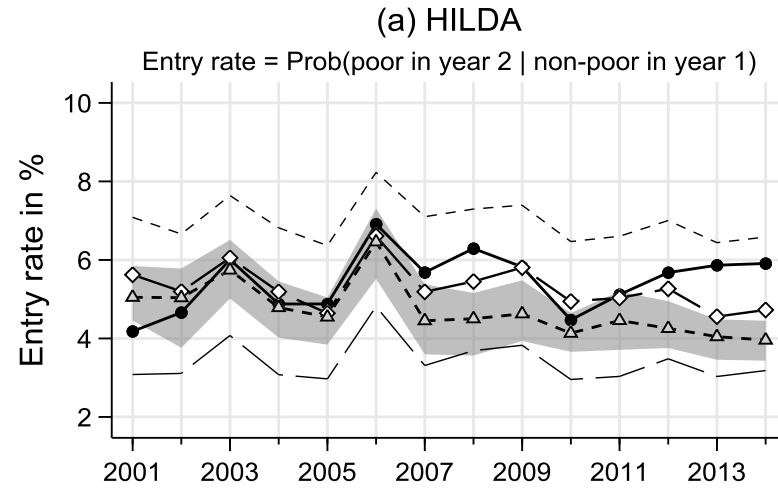
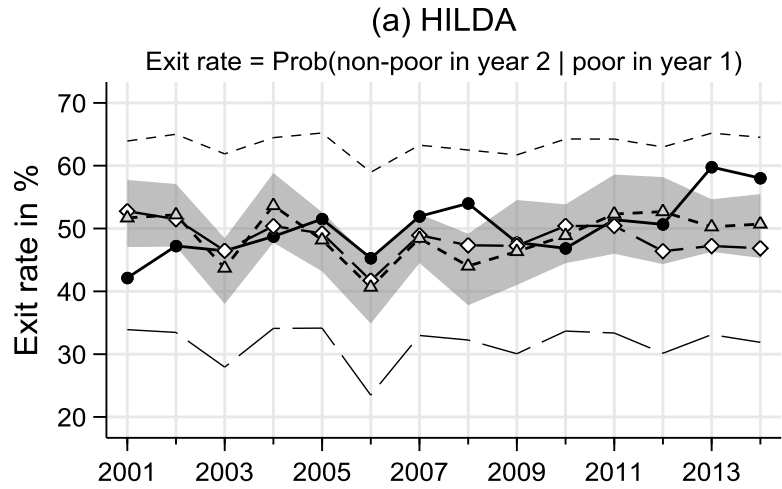


Figure 10. Estimates of poverty exit and entry rates, by year, individuals aged 18–59 ('leading case' definitions otherwise)

