

EU-SILC and the potential for synthetic panel estimates

Brian Colgan

October 15, 2020

Abstract

In the absence of panel data, researchers have devised alternative methods for estimating synthetic poverty dynamics using repeated cross-section surveys. These methods are not only salient in the absence of panel data, but also in contexts where there are concerns over the quality of panel data and/or the panel data is of insufficient length to analyze medium to long term mobility trends. Both of these issues afflict the longitudinal element of the European Survey on Income and Living Conditions (EU-SILC) (Hérault and Jenkins, 2019). It is the aim of this paper to assess the accuracy of one such synthetic panel approach in the context of EU-SILC.

Dang et al., 2011 (henceforth DLLM) and Dang and Lanjouw, 2016 (henceforth DL) outline an approach for producing synthetic panel estimates. DLLM sets out a procedure for estimating parametric and non-parametric bounds while DL presents an innovation which is capable of producing point estimates of poverty dynamics. Both approaches link independent households over time via income models containing only time invariant characteristics. The residuals from these income models are assumed to follow a bivariate normal distribution, which under certain assumptions concerning the correlation of residuals, allows parametric bounds to be estimated. The DL innovation proposes the use of pseudo panel techniques to approximate the correlation in residuals over time thus allowing for the estimation of point estimates of poverty dynamics. Previous validation exercises have provided support for the accuracy of the DLLM and DL approaches (Dang and Lanjouw, 2016 and Garces Urzainqui, 2017) while Hérault and Jenkins, 2019, the largest validation effort to date, are much less optimistic and have raised concerns over the accuracy of the DL innovation.

This paper utilizes the EU-SILC longitudinal data for France, Poland and Greece to further assess the performance of DLLM and DL synthetic poverty dynamics. For each country a series of one year panels are constructed for the period 2005-2016. In order to avoid results being driven by the same households being present in both the longitudinal dataset and the hypothetical cross-sectional data, the longitudinal data is randomly split in half; one half of the data is used for the income model in period 1 while the other half is used for the income model in period 2. This process is repeated R times to avoid spurious results relating to any particular split of the data.

The accuracy of synthetic panel estimates is found to be highly sensitive to the normality of residuals from the income models. If the residuals from the income model deviate greatly from the normal distribution then the bivariate normality assumption will not hold and the estimates produced will be inaccurate. Excluding outliers and transforming the data are found to have a significant impact on accuracy. When the true correlation is known and steps have been taken to improve the normality of residuals, the synthetic panel approach produces highly accurate estimates of poverty dynamics at the aggregate and sub-population levels. The majority of synthetic panel estimates lie within or marginally outside the 95 percent confidence interval of the true panel estimate when the poverty line is set at 60 percent of the median. Similar to Hérault and Jenkins, 2019 the DL innovation is found to be highly sensitive to cohort definition. Alternatives to the DL innovation such as using the correlation term from neighboring countries with panel data and/or using short run longitudinal data to estimate longer term synthetic panel estimates are explored. These alternative approximates are found to be quite accurate.

The well documented shortcomings of the longitudinal element of EU-SILC means there is great potential for synthetic panel approaches to not only overcome problems of attrition, but also to produce medium and long run estimates of poverty dynamics and to incorporate EU-SILC's ad-hoc modules into the analysis of poverty dynamics. This paper provides supporting evidence as to the viability of the DLLM approach for EU-SILC as well as practical guidance for its implementation.

Keywords: Synthetic panel, pseudo panel, poverty, poverty dynamics, EU-SILC

JEL Codes: C53, D31, I32, O15

References

- Dang, H.-A., & Lanjouw, P. (2016). Measuring poverty dynamics with synthetic panels based on repeated cross-sections.
- Dang, H.-A., Lanjouw, P., Luoto, J., & McKenzie, D. (2011). *Using repeated cross-sections to explore movements in and out of poverty*. World Bank.
- Garces Urzainqui, D. (2017). Poverty transitions without panel data? an appraisal of synthetic panel methods.
- Hérault, N., & Jenkins, S. P. (2019). How valid are synthetic panel estimates of poverty dynamics? *The Journal of Economic Inequality*, 17(1), 51–76.