Application of 'pseudo panels' to investigate causal link between HIV and fertility in sub-Saharan Africa

M.A. Magadi, School of Social Sciences, University of Hull.

Abstract

Panel data are valuable for answering questions about change over time, but remain relatively scarce in most developing countries, especially sub-Saharan Africa (SSA). Where there exists a series of repeated cross-sectional data, 'pseudo-panels' provide a promising alternative. The use of 'pseudo panels' has received considerable attention in econometrics, but application in Demography remains rare. This paper explores the potential for using 'pseudo panels' to investigate causal link between HIV and fertility in SSA. The relationship between HIV and fertility is a complex one, partly because causality can run in either direction. We focus primarily on fertility as the outcomes of interest and HIV as a contributing factor. Repeated cross-sectional Demographic and Health survey (DHS) data from 20 countries in SSA are used to construct "pseudo panels" based on birth cohorts by country. The pseudo panels allow an investigation of possible causal link between HIV in an earlier survey and fertility behaviour of similar cohorts in a subsequent survey. Measures of HIV are based on risk perception and HIV status, while fertility is based on births within the last five years and future fertility intention. A total of 140 cohorts (7 age groups * 20 countries) were constructed, with 120 cohorts having data for at least two time periods. The analysis used two alternative approaches: (i) 'Conditional' models of HIV in an earlier survey and fertility behaviour in subsequent surveys; and (ii) Repeated measures multilevel analysis, with cohort as Level-2, and measurement occasion as level-1. An evaluation /assessment of the analysis involved a comparison of findings from 'pseudo cohort' with individual-level analysis, and a multi-level estimation of intra-cohort correlation coefficients to assess the degree of similarity of individuals in the same cohort. A multivariate analysis based on fixed effects models was used to determine the extent to which observed patterns may be attributable to key demographic/ socio-economic differences or infer possible causal links. Preliminary analysis shows promising results on application of 'pseudo panels' in investigation of demographic causal links in settings with limited panel data such as sub-Saharan Africa. However, further analysis is necessary for conclusive results. In particular, advanced modelling using Multiprocess modelling or Structural equation modelling will be used to address possible endogeneity in the relationships observed.

Background context

Panel data are valuable for answering questions about change over time, but remain relatively scarce in most developing countries, especially sub-Saharan Africa (SSA). Where there exists a series of repeated cross-sectional data, 'pseudo-panels' provide a promising alternative. In the 'pseudo panel' approach "similar" individuals are grouped in cohorts, and aggregate measures within these cohorts are treated as observations in a synthetic panel (Deaton, 1985). The use of 'pseudo panels' has received considerable attention in econometrics (Browning, Deaton and Irish, 1985;Verbeek and Nijman 1992; Moffitt, 1993; Collado, 1997; Propper, Rees and Green, 2001; McKenzie, 2004; Verbeek, 2008), but application in Demography remains rare. This paper explores the potential for using 'pseudo panels' to investigate causal link between HIV and fertility in SSA. The relationship between HIV and fertility is a complex one, partly because causality can run in either direction. We focus primarily on fertility as the outcomes of interest and HIV as a contributing factor.

Data and Methods

Repeated cross-sectional Demographic and Health survey (DHS) data from 20 countries in SSA are used to construct "pseudo panels" based on birth cohorts by country. The pseudo panels allow an investigation of possible causal link between HIV in an earlier survey and fertility behaviour of similar cohorts in a subsequent survey. The 20 countries included in the analysis represent all countries in sub-Saharan Africa where the DHS has included HIV testing on nationally representative samples of women of reproductive age in at least two consecutive surveys (see Table 1). Measures of HIV are based on risk perception and HIV status, while fertility is based on births within the last five years and future fertility intention.

A total of 140 cohorts (7 age groups * 20 countries) were constructed, with 120 cohorts having data for at least two time periods. The analysis used two alternative approaches: (i) 'Conditional' models of HIV in an earlier survey and fertility behaviour in subsequent surveys; and (ii) Repeated measures multilevel analysis, with cohort as Level-2, and measurement occasion as level-1. An evaluation /assessment of the analysis involved a comparison of findings from 'pseudo cohort' with individual-level analysis, and a multi-level estimation of intra-cohort correlation coefficients to assess the degree of similarity of individuals in the same cohort. A multivariate analysis based on fixed effects models was used to determine the extent to which observed patterns may be attributable to key demographic/ socio-economic differences or infer possible causal links.

Table 1: Summary of DHS (and AIS[1]) data in the study

	Country /Year of survey	Women tested for HIV	% HIV+	Number of cases HIV+
1	*Burkina Faso (2003, **2010**)	8384	1.2	101
2	*Cameroon (2004, **2011**)	7221	5.6	404
3	*DR Congo (2007, **2013/14**)	9264	1.6	148
4	*Cote d'Ivoire (2005**, 2012**)	4509	4.6	207
5	*Ethiopia (2005, **2011**)	5942	1.9	113
6	*Guinea (2005, **2012**)	4622	2.1	97
7	*Kenya (2003, **2008/9;** AIS: 2007, 2012)	3641	8.0	291
8	*Liberia (2007, **2013**)	4397	2.4	106
9	*Lesotho (2004, **2009**)	3800	27.0	1026
10	*Malawi (2004, **2010,** 2015-16)	7091	12.9	915
11	*Mali (2006, **2013**)	4806	1.3	62
12	*Mozambique (**2009**)	6000	13.1	786
13	*Niger (2006, **2012**)	5000	0.4	20
14	*Rwanda (2005, **2010**)	6917	3.7	256
15	*Senegal (2005, **2011**)	5300	0.8	42
16	*Siera Leone (2008, **2013**)	7695	1.7	131
17	*Tanzania (**2010**,AIS:2003, 2007, 2011/2)	8179	6.6	540
18	*Uganda (2004/05, **2011**)	11950	8.3	992
19	*Zambia (2007, 2**013**)	14719	15.1	2223
20	*Zimbabwe (2005/6, **2010/11**, 2015)	7313	18.0	1316
	Overall sample	179747		13071

*Data with HIV test results for these countries are available for multiple rounds of DHS surveys, enabling construction of 'pseudo-panels' for causal analysis. Data shown (marked in **bold**) are for the latest survey, unless the latest survey is still on-going and data not yet available**.**
1 – AIDS Indicator Survey

Preliminary findings

Preliminary findings suggest that perceived risk has a stronger correlation with fertility behavior and intentions, than HIV status. Observed patterns suggest that in settings where a large proportion of individuals do not know their HIV status, perceived risk of HIV is a stronger determinant of fertility behavior than the actual HIV status. Stronger associations were observed between HIV and fertility intentions than actual fertility, consistent with patterns observed at individual level. It is interesting to note that for HIV status, the difference in fertility intention between HIV-positive and HIV-negative women is notably larger in later than in earlier survey, presumably since a higher proportion of individuals in the later surveys knew their HIV status. For perceived risk, the differences between none/low and moderate/high risk is more or less the same at different time periods.

Discussion and further analysis

The preliminary analysis shows promising results on application of 'pseudo panels' in investigation of demographic causal links in settings with limited panel data such as sub-Saharan Africa. Indeed, 'pseudo panels' do have some merits over true panels since such data are immune to attrition bias or other threats to internal validity such as testing bias or design contamination. However, they are prone to instrumentation bias if the study variables were measured differently in subsequent surveys. Furthermore, an important consideration in this application is whether the number of DHS rounds considered here (typically two) is adequate. Nevertheless, it has been noted that 'In longitudinal repeated measures design, we usually have a large number of level 2 units with rather few Level-1 units (Rasbash et al, 2015: 195), and an advantage of multilevel modelling of repeated data is the ability to handle unequal measurement intervals (Rasbash et al, 2015).

Although overall preliminary results seem plausible, further analysis is necessary for conclusive results. First, it is important to establish whether it is appropriate to lag or not to lag the dependent variable. More importantly, advanced modelling is necessary to address possible endogeneity in the relationships observed. This will be undertaken using Multiprocess modelling or Structural equation modelling.

References

Browning M, Deaton A, Irish M (1985). A profitable approach to labor supply and commodity demands over the life cycle. *Econometrica* 53:503-543

Collado, M.D. (1997), Estimating Dynamic Models from Time Series of Independent Cross-Sections, *Journal of Econometrics*, 82, 37–62.

Deaton, A (1985). Panel data from time series of cross-sections. *Journal of Econometrics* vol. 30(1-2): 109-126.

McKenzie, D.J. (2004), Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels, *Journal of Econometrics,* 120, 235–262.

Moffitt, R. (1993), Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections, *Journal of Econometrics*, 59, 99–123.

Propper, C., H. Rees and K. Green (2001), The Demand for Private Medical Insurance in the UK: A Cohort Analysis, *The Economic Journal*, 111, C180–C200.

Rasbash, J., Steele, F., Browne, W.J. and Goldstein, H. (2015) *A User's Guide to MLwiN, v2.33.* Centre for Multilevel Modelling, University of Bristol.

Verbeek, M (2008). Pseudo-panels and repeated cross-sections in Matyas, L. and Sevestre, P (Eds). *The Econometrics of Panel Data: Fundamentals and recent developments in theory and practice.* Advanced Studies in theoretical and Applied econometrics, Third Edition, Volume 46: 369-383 (Chapter 11), Springer.

Verbeek M. and Nijman T (1992). Can Cohort Data be Treated as Genuine Panel Data? *Empirical Economics* 17:9-23