it has played a considerable role. It may be expected that the direct effects of the decrease in research funds will be more evident in the subsequent years. The observed decline in research productivity in 2010–11 may also be a result of the considerable psychological stress that accompanies a financial crisis and also affects researchers. In conclusion, a decline in relative biomedical research productivity in Greece was observed, starting shortly after the initiation of the financial crisis, the precise contribution of which, however, is unclear.

References


doi:10.1093/ije/dys087
Advance Access publication 2 July 2012
© The Author 2012; all rights reserved.

Alternatives to principal components analysis to derive asset-based indices to measure socio-economic position in low- and middle-income countries: the case for multiple correspondence analysis

From PIERRE TRAISSAC* and YVES MARTIN-PREVEL

IRD (Institut de Recherche pour le Développement), NUTRIPASS Research Unit, IRD-UM2-UM1, 911, av. Agropolis Montpellier, France

*Corresponding author. IRD (Institut de Recherche pour le Développement), NUTRIPASS Research Unit, IRD-UM2-UM1, 911, av. Agropolis Montpellier, France. E-mail: pierre.traissac@ird.fr

We would like to congratulate and thank Howe et al. for their recent methodological article on measuring socio-economic position for epidemiological studies in low- and middle-income countries. Indeed, their conceptual framework for analysis and measurement of the different dimensions of socio-economic position is of major interest for epidemiological researchers, whether for the purpose of assessment of socio-economic status as a main exposure under study or to control for confounding. Nevertheless, we would like to add a few comments regarding the computation of asset-based household welfare indices using principal components analysis (PCA).

First, although PCA is the most often used method for deriving data-based weights for computation of asset-based indices from different household asset items, it is essentially an exploratory data analysis method; therefore, its appropriateness could be discussed in the framework of ‘exploratory’ vs ‘model-based’ factor analysis methods based on linear correlation coefficients, as, indeed, the objective of asset-based estimation of household welfare is to estimate a latent ‘welfare variable’ from the manifest household assets, according to structural equation models terminology. 2

A different point is that use of PCA for analysis of the most often categorical asset items (which are not always only binary) is somewhat problematic; this is acknowledged by the authors who refer to technical bypasses, such as computing tetrachoric correlation coefficients before analysis and weighing of the items by PCA to deal with the issue. To our point of view, an exploratory data analysis method, such as multiple correspondence analysis (MCA), 3 could be more appropriate for the purpose of deriving asset-based indices in low- and middle-income countries. The authors do make an indirect reference to such a method that they briefly discussed in a previous work, 4 but where they argued that MCA is not simpler to implement or understand than PCA and cannot be used with a mixture of variables of different types. But, indeed, we think that MCA should all the more be acknowledged as an alternative to PCA for deriving asset-based scores, as, although it is initially designed specifically for categorical variables, it is the only such multivariate method that can be used to analyse any mixture of binary, categorical, discrete or continuous variables (provided the latter are suitably categorized). 5 Thus, it can be used for various purposes, either for analysis of categorized interval variables without any of the distributional or linearity assumptions on which correlation coefficients rely (e.g. for deriving dietary patterns from food consumption data 6,7) or for deriving asset-based measures of household scores from a mix of interval and categorical variables (not restricted to binary) in various contexts 8–11 or a variety of other purposes. 12 In addition, as it is based on the same mathematical foundations of eigenvectors and eigenvalues of a matrix, someone
with a basic understanding of PCA can easily interpret MCA results. However, given the current trend of using more complex data analysis and modelling methods for better analysis of epidemiological data, we are also not entirely sure that simplicity is per se a relevant criterion. Also, for most MCA computer programs, graphical outputs, such as projections of categories on principal components, plots of categories coordinates and so on, are most helpful for interpretation of derived scores and weighing schemes. Moreover, for deriving household’s welfare indices from a set of categorical items, it could even be that using MCA is simpler than PCA, as it does not imply any previous computations as those quoted by the authors to ‘squeeze’ categorical data in the framework of PCA analysis (based on linear correlation coefficients). Also, MCA and PCA are currently available in most statistical packages used to analyse data from epidemiological studies: e.g. mca commands in Stata (Stata Corp LP, College Station, Texas, USA) and R (R Foundation for Statistical Computing, Vienna, Austria), Proc Corresp procedure in SAS (SAS Institute Inc., Cary, North Carolina, USA). Therefore, regarding measurement of socio-economic position, we do think MCA should be, if not straightforwardly preferred to, at least acknowledged as a serious alternative to PCA to derive asset-based household welfare indices. Generally, MCA is a multi-purpose multivariate analysis method with many potential uses in the analysis of epidemiological studies, and we think it should be more widely acknowledged beyond some specific domains (ecology, social sciences) where it is more mainstream.

Finally, we would like to underline that whatever the method, the apparent simplicity of computer implementation, as acknowledged by the authors for PCA, must not preclude thoughtful assessment of the underlying principles and also, most importantly, careful choice of the items and their categorizations. Beyond the points discussed above, that we hope will be useful to the readers, we would like to thank again Howe et al. for their major article and for sharing with us their in-depth understanding and analysis of measurement of socio-economic position in the specific context of low- and middle-income countries.

References