Productivity and education: Some comments to Bogetoft and Wittrup

Timo Kuosmanen[1]

1. Introduction

There is a large and growing stream of literature on productivity of educational institutions that overlaps such disciplines as educational economics, educational statistics, pedagogy, and operations research. Simplest thinkable productivity measure of a school is the pupil-teacher ratio, which if frequently referred to in public discussion. However, such simple ratios overlook such factors as the quality of educational outcomes, family background of pupils, economic resources of the school and its operating environment.

Bogetoft and Wittrup (2011, in this issue) approach this topic from a novel perspective that combines *multi-level modeling* (MLM) from the educational research (e.g., Raudenbush and Bryk, 2002) with *data envelopment analysis* (DEA, Charnes et al., 1978) used in applied economics and operations research. This allows the authors to take into account both qualitative and quantitative factors suggested above. I find this a very useful contribution towards further integration of this scattered field of research that is too much divided by disciplinary boundaries. However, I also see scope for a more systematic and rigorous integration of the methods. Purpose of this note is to provide some thoughts for further development of this line of research. I start from methodological issues and proceeding via scale efficiency to practical implementation.

2. Further integration of methodology

MLM is mainly confined to educational research, but it is analogous to the standard *fixed effects* (FE) and *random effects* (RE) models in econometrics (e.g., Greene, 2012). While the FE and RE models are usually applied to panel data where n schools are observed over T time periods, one could equally well apply them to a cross-section of n schools that are observed from the perspectives of T_i individual pupils (assume an unbalanced panel where T_i differs across schools). The fixed effects of the FE model can be used for estimating the "value added" of the school, analogous to the MLM approach. Interestingly, in the literature of *stochastic frontier analysis* (SFA, e.g., Kumbhakar and Lovell, 2003), the econometric panel data methods are similarly used for estimating productivity differences across schools. Such SFA models could readily incorporate both pupil-level and school-level information in a coherent and fully-integrated framework.

The main advantage of DEA to SFA is that it does not require any particular functional form for the production frontier. While DEA originated as a mathematical programming approach, the statistical properties of the DEA estimator have been established (Banker, 1993; Simar and Wilson, 2000). Recently, Kuosmanen and Johnson (2010) have shown that DEA is a sign-constrained variant of convex regression (Hanson and Pledger, 1976). This link between DEA and regression analysis enables us to integrate a DEA-style nonparametric frontier to the econometric models of panel data (see Kuosmanen and Kortelainen, 2011, for details). A practical challenge with such an integrated modeling approach is the enormous computational burden when the sample size (number of pupils) is very large. However, more efficient computational algorithms are being developed. Moreover, clever use of modeling can significantly reduce the computational burden.

The main disadvantage of DEA is that it does not take into account an explicit disturbance term, and hence any measurement errors, omitted variables and other noise is attributed to inefficiency. Incorporating DEA-style axiomatic frontier in an econometric framework as in Kuosmanen and Kortelainen (2011) can effectively solve this problem.

Finally, the two-stage method where DEA efficiency scores are regressed on environmental variables z (which Bogetoft and Wittrup consider at the end of Section 5) has been sharply criticized by Simar and Wilson (2007), who argue that the statistical inferences in the second-stage regression are invalid. Integrating the DEA-style nonparametric frontier to the econometric model solves this problem as well (see Johnson and Kuosmanen, 2012).

2. Scale efficiency

An important issue from the policy perspective is the optimal school size, which is closely related to the optimal pupil-teacher ratio: small schools tend to have smaller groups than large schools. Small groups obviously require more teacher resources than large groups, but there seems to be a tradeoff between the

[1] School of Economics, Aalto University, Helsinki, Finland. timo.kuosmanen@aalto.fi.

group size and the quality of educational outcomes. Further, if small schools are closed and pupils are reallocated to larger units, the transportation costs can increase significantly, especially in sparsely populated areas. Estimating the economically optimal school size is a challenging task that involves several tradeoffs.

Bogetoft and Wittrup address this issue using the DEA-based scale efficiency measure. From conceptual point of view, scale efficiency is perfectly appropriate for evaluating the most productive school size. However, the input variables considered in this study restrict to labor inputs (teacher and staff hours). If data permits, the scale efficiency analysis could easily be extended to include the capital stock and other relevant inputs. However, to estimate the potential cost saving obtainable by scale efficient operation of schools from the social point of view, one should look beyond the school-specific input resources and take into account external costs (e.g., transportation cost) that are not included in the school budget.

3. Realizing the cost saving potential

Achieving the potential cost saving estimated by frontier methods is challenging in practice. In this respect, Bogetoft and Wittrup devote considerable attention to interactive benchmarking where the inefficient schools can learn from their efficient peers. Benchmarking is widely recognized as a useful strategy to facilitate diffusion of best practices. However, benchmarking does involve an opportunity cost (consider the time devoted for searching the appropriate peers, visiting the schools and organizing meetings concerning school practices, etc.), whereas the benefits are uncertain. It is not always self-evident that the benefits outweigh the cost. To my knowledge, there is little evidence that DEA-based benchmarking is more cost-effective than some simpler and cheaper alternative for diffusion of best practices. I would like to suggest this as an interesting topic for further research.

The approach that I would propose is to first identify good practices that schools have (e.g., pedagogical approach, extracurricular activities, the use of ICT), then collect data regarding the practices, and finally apply efficiency analysis to assess how different practices affect performance. The effects of school practices (or local or national policy) on productivity could be estimated similar to the *z*-variables that represent the operating environment (see Johnson and Kuosmanen, 2012, for a more detailed discussion). Results of such estimation could provide valuable information for both educators and policy makers about the productivity improving practices.

References

Banker, R.D. (1993) Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39: 1265-1273.

Bogetoft, P., Wittrup, J. (2011) Productivity and Education: Benchmarking of elementary and lower secondary schools in Denmark, *Nordic Economic Policy Review*, this issue.

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research* 2: 429–444.

Greene, W.H. (2012). *Econometric Analysis*, 7th Edition, Prentice Hall.

Johnson, A.L., Kuosmanen, T. (2012) One-Stage Estimation of the Effects of Operational Conditions and Practices on Productive Performance: Asymptotically Normal and Efficient, Root-*n* Consistent StoNEZD method, *Journal of Productivity Analysis*, in press, DOI: 10.1007/s11123-011-0231-5.

Kumbhakar, S, Lovell, C.A.K. (2003) *Stochastic Frontier Analysis*, Cambridge University Press.

Kuosmanen, T. Johnson, A.L. (2010) Data Envelopment Analysis as Nonparametric Least Squares Regression, *Operations Research* 58(1): 149-160.

Kuosmanen, T., Kortelainen, M. (2011) Stochastic Non-Smooth Envelopment of Data: Semi-Parametric Frontier Estimation Subject to Shape Constraints, *Journal of Productivity Analysis*, in press, DOI: 10.1007/s11123-010-0201-3.

Raudenbush, S.W., Bryk, A.S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Method* (2nd Edition). Newbury Park, CA: Sage.

Simar, L., Wilson P.W. (2000) Statistical Inference in Nonparametric Frontier Models: The State of the Art , *Journal of Productivity Analysis* 13: 49-78.

Simar, L., Wilson P.W. (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136(1): 31-64.