

Scale and Scope Economies, Higher Education

Jill Johnes, Huddersfield Business School, University of Huddersfield

Economies of scale occur if the average (or unit) cost of producing an output falls as the volume of production expands. Universities can be seen to be producing (in their simplest form) graduates, and the unit cost of producing each graduate is likely to fall as student numbers rise, as they are taught in ever larger classes. These economies of scale are unlikely to continue indefinitely: as the organisation becomes larger, complexity and bureaucracy from increasing processes to cope with higher production also rise to a point where the additional costs exceed any fall in unit costs of production from greater size. At some level of production, then, diseconomies of scale will set in. Economies of scale can be identified in the single-product case by examining the (long run) average and marginal cost curves. If the ratio of the former to the latter is greater (less) than unity, for a given production level, there are economies (diseconomies) of scale at that point.

The assumption above that universities only produce graduates is somewhat simplistic. In fact, universities are complex organisations producing multiple outputs (from multiple inputs). For example, undergraduates graduate from a variety of different disciplines, there are graduates from Masters and research programmes, there are research outputs such as publications and patents, and there are outputs relating to third mission activities arising from engagement with businesses and the wider community. The notion of economies of scale in the single output case has been extended to the multi-product context (Baumol *et al.* 1982), and can be applied in the higher education setting. Thus ray economies (diseconomies) of scale are the cost savings (or dissavings) which occur when all outputs increase (holding the output mix constant). Product-specific economies (diseconomies) of scale are the savings (dissavings) in costs occurring when one output increases and all other outputs remain at fixed production levels (Johnes *et al.* 2008).

Economies of scale are distinct from economies of scope which occur when it is less costly to produce a range of outputs together than to produce each output separately in its own production unit. In the university context, there are likely to be savings from producing together graduates from across a variety of disciplines, or from producing teaching and research together, rather than having separate organisations each focusing on a specific discipline (for example, arts, sciences, medicine and dentistry), or each producing teaching only or research only. Two specific concepts can be defined. Global economies of scope occur when the costs of producing all outputs together in a single firm are less than the sum of the costs of producing each output in a separate firm. Product-specific economies of scope for product j arise when the costs of producing all outputs together in a single firm are less than the sum of costs of producing output j in a separate firm and all outputs apart from j in another firm (Johnes *et al.* 2008).

Economies of scope can arise if it is possible to spread the costs of central services across an array of outputs. In higher education, these services might be marketing, placement and careers services, research support, and financial services. They might also arise because the outputs produced are joint products: research and teaching, for example, might be complementary activities such that undertaking research feeds into teaching/supervision, and vice versa.

In order to identify empirically whether or not economies of scale and scope exist in a higher education sector typically requires specification and estimation of a cost function using

appropriate estimating techniques. A cost function describes the costs in relation to the level of output produced by the institution, and information about the price or quality of the institution's inputs (Johnes *et al.* 2005). The precise choice of functional form for the cost equation should ideally conform to various criteria identified by Baumol *et al.* (1982), namely that the functional form should:

- Be consistent with cost minimisation given outputs and input costs. This means (amongst other things) that it must be a non-negative and non-decreasing function.
- Provide sensible predictions of costs where output of one or more products is zero. Thus functions which include logarithmic transformation of outputs such as the Cobb-Douglas or the translog cost functions are not appropriate. This criterion is essential for estimating economies of scale and scope.
- Neither prevent nor impose the existence of scale or scope economies or diseconomies. A linear function, for example, precludes economies of scope and allows only limited economies of scale.

Functional forms such as the cross elasticity of substitution, quadratic, and hybrid translog fulfil these criteria, they have various advantages and disadvantages in terms of estimation, and have been used in previous empirical studies.

One challenge is that identifying the presence of economies of scale requires the estimation of a long run cost function, and this assumes that in a given time period the HEIs under study are operating on their long run cost curve. Yet the data (costs by HEI) observed in a given time period are likely to be a mix of observations: some HEIs will be operating on their long run cost curve; some will be on a short run cost curve prior to transition to the long run position; some may be in a long run position but may not be operating efficiently. It is difficult for the analyst to recognise which HEIs are operating on short run cost curves and which on long run cost curves, particularly if the period under study is one of rapid change, in which case HEIs may be in a state of adjustment. It is also challenging to know which of the observed HEIs are cost efficient.

For these reasons, it has become customary to estimate the long run cost function by means of a frontier estimation technique since this results in an estimated function which envelops the data. This contrasts with an ordinary least squares regression approach (or related method) which estimates an average line through the observed points. With a frontier estimation technique, the position of the cost function is determined by those observations which are likely to be both efficient and in a long run equilibrium position. The derivation of efficiency using a frontier estimation approach is rooted in the work of (Farrell 1957). The frontier techniques can be parametric, such as the family of estimation methods falling under the heading of stochastic frontier analysis (SFA) (Aigner *et al.* 1977), or non-parametric, such as those falling under the data envelopment analysis (DEA) umbrella (Charnes *et al.* 1978;1979).

A basic SFA approach estimates identical cost function parameters for all HEIs in the sector, permits statistical inference, and allows calculation of economies of scale and scope from the estimated parameters – these can be, for example, HEI-specific, or average across the sector, or predictions for large or small HEIs (such as double, treble or half the observed average size of HEI). Increasingly, though, researchers are taking into account the fact that different HEIs, or different *groups* of HEIs, might transform their inputs into outputs differently. If such assumptions are made, then DEA, underpinned by linear programming methods, is an ideal choice of estimation method as it permits input and output weights to differ across institutions.

It is, however, difficult to derive estimates of economies of scale and scope when DEA is used to evaluate the cost technology. Recent developments in SFA, such as random parameters SFA (Tsonas 2002; Greene 2005), where parameters vary by each individual HEI, or latent class SFA (Lazarsfeld and Henry 1968; Orea and Kumbhakar 2004), where parameters are derived for groups of HEIs (where the groups are determined by the data), permits both variation in parameters across (groups of) HEIs *and* calculation of economies of scale and scope.

There is now a considerable literature concerning the cost structure of HEIs. An early study of university cost functions (Verry and Layard 1975) accepts that universities produce both teaching and research, but the linear cost function is inadequate in terms of modelling economies of scale and scope (Johnes and Johnes 2016). The seminal breakthrough in estimating higher education cost functions came with the work of Cohn *et al.* (1989), which was the first to incorporate the concepts of Baumol *et al.* (1982) into the higher education context, specifically estimating economies of scale and scope for both public and private HEIs in the United States. Subsequent studies followed focusing on various higher education sectors (see, for example, de Groot *et al.* 1991; Dundar and Lewis 1995; Glass *et al.* 1995; Koshal and Koshal 2001; Sav 2011; Worthington and Higgs 2011). These studies were largely unanimous in their findings of ray economies of scale, but results for economies of scope were mixed. These findings, however, were all based on an estimation approach which failed to take into account the presence or otherwise of inefficiency in the sector.

Application of frontier estimation methods combined with a multi-product cost function specification has now become the norm (for a review, see Johnes and Johnes 2016). Findings across various higher education sectors based on frontier estimation suggest that ray economies of scale are typically exhausted, but there are some product-specific economies relating to research and/or production of postgraduate outputs. Studies employing both frontier and panel data estimation methods (allowing for unobserved heterogeneity amongst the HEIs) find evidence of diseconomies of scope both globally and for individual products. Thus studies applying frontier estimation methods generally find lower levels of economies of scale and scope than those using non-frontier methods.

There are some caveats around these findings. First, the choice of estimation method and functional form can affect the conclusions regarding economies of scale and scope. Application of frontier rather than non-frontier estimation methods negatively impacts estimates of economies of scale and scope, as does the adoption of a quadratic cost functional form (Zhang and Worthington 2018). Researchers should recognise these effects when choosing their estimation method and functional form, and make their choices based on the underlying theory.

The second caveat relates to the issue of third mission activities. While researchers can typically find adequate measures of the teaching and research outputs to include in their cost function, it is much more difficult to find acceptable measures of third mission activities, and so this output is ignored in many studies. This omission may bias estimates of economies of scale and scope.

A third caveat concerns diversity. HEIs can differ from each other in terms of, for example, mission, size, age, subject mix, and research intensity. The estimated cost function (and hence degree of economies of scale or scope) can vary depending on these characteristics. Such diversity has been addressed in empirical studies using a variety of possible approaches such as including a binary variable in the functions to reflect mission group, separate estimation of cost function by pre-defined mission groups, or adding exogenous control variables to the cost

equation. Most recently, researchers have adopted random parameter frontier estimation approaches (see, for example, Johnes and Johnes 2009; Agasisti and Johnes 2010;2015), and have typically found that, on average, ray economies of scale are exhausted or decreasing, although there are some product-specific economies (the precise nature of which varies by country), while economies of scope are typically negative. Studies which employ latent class frontier estimation models are relatively rare (see, for example, Johnes and Johnes 2013; Johnes and Johnes 2016), and find that the presence or otherwise of economies of scale and scope varies by the group of institutions.

It is clear that findings on economies of scale and scope can vary by type of HEI. It therefore seems that there is an opportunity to undertake a more in-depth analysis of why some HEIs have scale or scope advantages and why some do not (Sav 2011; Hemelt *et al.* 2018). Understanding the mechanisms by which, or the context in which, universities can achieve economies of scale or scope seems to be a topic which has received less attention than the actual empirical estimation of these concepts, but it is one where further exploration would provide useful policy and managerial insights.

References

- Agasisti, T. and G. Johnes (2010). 'Heterogeneity and the evaluation of efficiency: The case of Italian universities.' *Applied Economics* 42(11): 1365-1376.
- Agasisti, T. and G. Johnes (2015). 'Efficiency, costs, rankings and heterogeneity: The case of US higher education.' *Studies in Higher Education* 40(1): 60-82.
- Aigner, D., C. A. K. Lovell and P. Schmidt (1977). 'Formulation and estimation of stochastic frontier production models.' *Journal of Econometrics* 6: 21-37.
- Baumol, W. J., J. C. Panzar and R. D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. London, Harcourt Brace Jovanovich.
- Charnes, A., W. W. Cooper and E. Rhodes (1978). 'Measuring the efficiency of decision making units.' *European Journal of Operational Research* 2(4): 429-444.
- Charnes, A., W. W. Cooper and E. Rhodes (1979). 'Measuring the efficiency of decision making units: a short communication.' *European Journal of Operational Research* 3(4): 339.
- Cohn, E., S. L. W. Rhine and M. C. Santos (1989). 'Institutions of higher education as multi-product firms: economies of scale and scope.' *Review of Economics and Statistics* 71(2): 284-290.
- de Groot, H., W. W. McMahon and J. F. Volkwein (1991). 'The cost structure of American research universities.' *Review of Economics and Statistics* 73(3): 424-431.
- Dundar, H. and D. R. Lewis (1995). 'Departmental productivity in American universities: Economies of scale and scope.' *Economics of Education Review* 14(2): 119-144.
- Farrell, M. (1957). 'The measurement of productive efficiency.' *Journal of the Royal Statistical Society Series A* 120(3): 253-281.
- Glass, J. C., D. G. McKillop and N. S. Hyndman (1995). 'The achievement of scale efficiency in UK universities: a multiple-input multiple-output analysis ' *Education Economics* 3(3): 249-263.
- Greene, W. (2005). 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model.' *Journal of Econometrics* 126: 269-303.
- Hemelt, S. W., K. M. Stange, F. Furquim, A. Simon and J. E. Sawyer (2018). 'Why is math cheaper than English? Understanding cost differences in higher education.' NBER Working Paper No. 25314.

- Johnes, G. and J. Johnes (2009). 'Higher education institutions' costs and efficiency: taking the decomposition a further step.' *Economics of Education Review* 28(1): 107-113.
- Johnes, G. and J. Johnes (2016). 'Costs, efficiency and economies of scale and scope in the English higher education sector.' *Oxford Review of Economic Policy* 32(4): 596-614.
- Johnes, G., J. Johnes and E. Thanassoulis (2008). 'An analysis of costs in institutions of higher education in England.' *Studies in Higher Education* 33(5): 527-549.
- Johnes, G., J. Johnes, E. Thanassoulis, P. Lenton and A. Emrouznejad (2005). *An Exploratory Analysis of the Cost Structure of Higher Education in England*. London, Department for Education and Skills, Research Report 641.
- Johnes, J. and G. Johnes (2013). *Efficiency in the Higher Education Sector: A Technical Exploration*. London, Department for Business Innovation and Skills.
- Koshal, R. K. and M. Koshal (2001). 'Multi-product total cost function for higher education: A case of bible colleges.' *Economics of Education Review* 20: 297-303.
- Lazarsfeld, P. F. and N. W. Henry (1968). *Latent Structure Analysis*. New York, Houghton Mifflin.
- Orea, L. and S. C. Kumbhakar (2004). 'Efficiency measurement using a latent class stochastic frontier model.' *Empirical Economics* 29(1): 169-183.
- Sav, G. T. (2011). 'Panel data estimates of public higher education scale and scope economies.' *Atlantic Economic Journal* 39(2): 143;153;-153.
- Tsionas, E. G. (2002). 'Stochastic frontier models with random coefficients.' *Journal of Applied Econometrics* 17: 127-147.
- Verry, D. W. and P. R. G. Layard (1975). 'Cost functions for university teaching and research.' *Economic Journal* 85: 55-74.
- Worthington, A. C. and H. Higgs (2011). 'Economies of scale and scope in Australian higher education.' *Higher Education* 61(4): 387-414.
- Zhang, L.-C. and A. C. Worthington (2018). 'Explaining estimated economies of scale and scope in higher education: A meta-regression analysis.' *Research in Higher Education* 59(2): 156-173.