

The "usefulness" of stochastic frontier analysis for health care

Gary D. Ferrier*

Department of Economics, University of Arkansas, USA

Received: 4 March 2014

Revised: 29 April 2014

Accepted: 30 April 2014

Abstract

Newhouse (1994) questioned the usefulness of frontier efficiency estimation for health care applications. In the intervening years, numerous modeling advances have improved the econometric approach to efficiency measurement known as stochastic frontier analysis (SFA) and myriad academic studies have analyzed efficiency of various health care providers using SFA. This article takes stock of the current "usefulness" of applications of SFA in the health care sector and offers suggestions on how this usefulness can be further enhanced.

Keywords: frontier efficiency; stochastic frontier analysis; health care efficiency

JEL Classification Codes: I12, D22, C51

1. Introduction

In developed countries, health care expenditures as a share of GDP have been rising for decades; hospital care is typically the largest category of health care spending. Thus, it should not be surprising that interest in health care efficiency, and hospital efficiency in particular, has been of interest to academics, health care providers, and policy makers for some time¹. In the mid-1970s, the study of efficiency was given a fresh set of tools with the introduction of econometric and linear programming methods to calculate technical efficiency relative to a "best-practice" production frontier. The econometric approach to quantifying technical efficiency measurement is often labeled "stochastic frontier analysis" (SFA), while the linear programming approach is commonly termed "data envelopment analysis" (DEA). Often viewed as competing methodologies (though they can also be viewed as complementary), SFA and DEA each have relative advantages and disadvantages vis-à-vis one another. These

*E-mail: gferrier@walton.uark.edu.

Citation: Ferrier, G.D. (2014) The "usefulness" of stochastic frontier analysis for health care, *Economics and Business Letters*, 3(1), 27-34.

¹ Improved efficiency of the health care system is among the "six aims for improvements" called for by the Institute of Medicine in the US. The other five aims are improved safety, effectiveness, patient-centeredness, timeliness, and equity (Institute of Medicine, 2001).

methods have since been extended to allow for measures of cost, allocative, profit, scale, and revenue efficiencies.

The earliest applications of SFA and DEA, respectively, to overall hospital performance of which I am aware are Wagstaff (1989) and Sherman (1984). Both frontier methods have been used in the academic literature to measure efficiency in health care often enough that Hollingsworth (2008) referred to the measurement of health care efficiency as a "small industry." As this "industry" grew, reviews of the literature followed—e.g., Hollingsworth et al. (1999), Hollingsworth (2003), Worthington (2004), Jacobs et al. (2006), Hollingsworth (2008), Rosko and Mutter (2008), Hussey et al. (2009), and Chilingirian and Sherman (2011).

While SFA and DEA have been widely applied to health care (e.g., hospitals, nursing homes, pharmacies, physician practices, cardiac units, health care systems, etc.) in the academic literature by researchers in a variety of disciplines, their use by health care decision makers has been limited. For example, with reference to the use of benchmarks determined by frontier techniques to guide reimbursement schemes, Lovell (2006, p. 5) wrote that, "Although frontier techniques have gained acceptance in other regulatory arenas, their potential remains unfulfilled in healthcare." With regard to the development of metrics to improve various dimension of health care, McGlynn et al. (2008, p. iii) lament, "But when it comes to measuring efficiency—one of the six domains of quality identified by the Institute of Medicine—we have seen much less light than heat." These remarks harken back to Newhouse's (1994) critique of frontier techniques as a useful tool for health economics. While methodological advances have occurred for both SFA and DEA since Newhouse published his critique, the need remains for additional progress in the application of these methods to the health care sector. McGlynn et al. (2008) suggested four criteria on which to evaluate the application of efficiency measurement to the health care sector; namely, are the measures (1) important; (2) scientifically sound; (3) feasible; and (4) actionable?

The purpose of this article is not to review the academic literature on SFA's application to health care—as noted above, this has been done. Therefore, references to work in this area will be kept to a minimum. Instead, this paper will briefly address how the "usefulness" of SFA as a tool for health care research can be enhanced.

2. The importance of efficiency measurement in health care

Clearly, efficiency is an important issue for a large and growing sector of most economies. The importance of the efficient provision of health care is evident from the interest in this topic shown by providers and policy makers. However, the importance of the academic literature to decision makers has been limited. Consider that while many potential determinants of efficiency have been examined, a strong consensus of the effects of these factors has not been reached. For example, Hollingsworth (2008) reports "cautious conclusions" that the efficiency of public hospitals is "potentially" greater than that of private hospitals "in certain settings" does not offer strong guidance for decision making. Even relatively confident knowledge of the overall technical efficiency of a hospital, for example, can be difficult to translate into policy—i.e., it would be difficult for a hospital to contract all inputs (or expand all outputs) proportionately, even in the long run.

Another factor that likely has limited the value of academic results to decision makers is the unit of analysis used in many studies—hospitals, nursing homes, etc. Even a small hospital is a complex entity. The first application of DEA in a health care setting (Jadlow and Wilson, 1982) examines the efficiency of nuclear medicine facilities; perhaps focusing on smaller, more homogeneous units of analysis would increase the value of academic analyses of efficiency. For example, rather than assess efficiency at the hospital level, Dismuke and Sena (1999) assess the efficiency of three commonly used diagnostic technologies—computerized

axial tomography (CAT), electrocardiogram (EEG), and echocardiogram (ECO)—within district and central hospitals in Portugal. Focus on such granular levels of activity would increase the homogeneity of units of analysis, mitigating aggregation and heterogeneity issues.

Many of the academic applications of SFA to health care have examined hospital efficiency. As noted above, hospitals represent the largest share of health care expenditures for most developed economies. However, expenditures on visits to physicians' offices and clinics and on medical goods (primarily pharmaceuticals) are also large; expenditures on long-term care are not as large, but are increasing. Relatively few academic studies examine the efficiency of these other health care functions. Of course, access to data may limit the applications of SFA.

Finally, one must consider how the "packaging" of the results in academic studies may be limiting their importance to decision makers. Enhanced packaging may include a better explanation of the confidence attached to efficiency measures, identification of "best-practice" units that could serve as role models (DEA provides this), and a more concerted effort to clarify the policy implications of findings (using SFA to study the results of "natural experiments" could help in this regard [e.g., changes in staffing requirements]). Of course, involving decision makers from industry and government in the creation of the "package" itself would increase the value of SFA applications in health care to those who could make use of the information.

3. Improving the scientific soundness of SFA applications in health care

Worthington (2004) notes that empirical efficiency analyses typically proceed in three steps—(1) specification of the econometric model of the stochastic frontier, (2) selection of the inputs and outputs to use in the analysis, and (3) a determination about how to examine the relationship of efficiency to various factors internal and external to the unit of analysis. Factors that affect the scientific soundness of SFA applications, some generic, some more specific to health care studies, for each of the three steps listed above will be considered in turn in this section. A common thread ties these steps together is the need to accommodate heterogeneity in production.

Specification of the parametric, stochastic frontier

Cost minimization is one of the fundamental concepts in the theory of the firm. But as Dor (1994, p. 329) notes, "By allowing firms to deviate from their optimal path for systematic reasons, frontier estimators in the general literature have helped to inject more realism into econometric modeling of firms, and now, of hospitals and nursing homes as well." The continued generalization of models can further "inject more realism" into studies of health care efficiency.

Given that inefficiency is a measure of distance from a frontier, the functional form of the frontier is an important consideration. The two most commonly used functional forms in health care applications of SFA are the translog and Cobb-Douglas (a special case of the translog). However, many other functional forms are available—e.g., the generalized quadratic Box-Cox, generalized Leontief, normalized quadratic, squared-root quadratic, constant elasticity of substitution, minflex translog, and minflex generalized Leontief. The empirical literature on health care efficiency demonstrates that the magnitudes of estimated (in)efficiency and the efficiency rankings of health care service providers depends on the parameterization of the stochastic frontier model, however, little explicit attention has been given to the sensitivity of efficiency scores to the choice of functional form. The broader use of alternative functional forms, and discussion of the sensitivity of results, seems warranted.

Another issue related to the parametric specification in SFA is the assumption that parameters are constant across all observations. In health care, regulation varies across political units, various ownership structures operate side by side in the same markets, case-mix and quality vary widely across providers, etc. Relaxing the assumption of common parameters across observations would help to account for heterogeneity across units of analysis.² To do so, stochastic frontiers can be estimated using either random parameters or latent class models. The former treats heterogeneity as a continuous function of the parameters to be estimated, while the latter assumes that observations belong to one of a finite number of groups, or classes, each having their own parameters. For example, Barros et al. (2011) use a latent class stochastic frontier to endogenously identify different segments in their sample of hospitals. They use a testing-down procedure based on likelihood ratio tests and information criteria to determine the number of classes in their sample. Identification of clusters within a sample should provide more targeted implications for decision makers.

In SFA, the error term is a composite of noise (typically assumed to be normally distributed) and inefficiency (one-sided errors). Distributional assumptions made on the inefficiency term include the truncated normal, half-normal (a special case of the truncated normal), exponential, and gamma distributions; empirical work is largely based on the first two distributions. The issue of robustness to different error distribution assumptions does not have an analytical solution; however, as with functional form, it would be useful to examine the consistency of empirical efficiency scores (and their rankings) under different distributional assumptions. Given the extensive empirical literature applying SFA to health care, there is relatively little research that examines the sensitivity of findings with respect to distributional assumptions on the inefficiency component of the error term in SFA models. In general, it is difficult to test distributional assumptions because only the composed error, the sum of the inefficiency and noise terms, is observed. Recently, however, approaches have been developed to test whether the distribution of the composed error is the distribution implied by the sum of a normally distributed noise component and a one-sided inefficiency component with the distribution specified in the model.

Finally, given that various sub-populations (e.g., public and private hospitals) often exist in the same market, that measurement error (see below) is likely an issue, etc., accounting for heteroscedasticity in the error term, and correctly specifying its form if present, is a potentially important issue when applying SFA to health care.

Specification of outputs and inputs

As with all service sectors, output in the health care sector is conceptually difficult to define (e.g., see Triplett, 2013). On the one hand, the "output" of this sector is "health." It is difficult to operationalize a measure of this "output;" moreover, the "production" of "health" is a complex process whose inputs include medical care, genetic and environmental factors, personal behaviors, etc. A more pragmatic approach is to view "medical care" as the output and capital, labor, supplies, etc. as the inputs into the production process. Even under this pragmatic approach, specification of outputs and inputs is difficult because health care providers, hospitals in particular, produce a broad spectrum of outputs of varying quality using a wide array of inputs which also differ in quality. In practice, SFA is applied using output and input measures that are readily available in existing data sets (e.g., the American Hospital Association's Annual Survey of Hospitals database). In many cases, the measures used are presented as "proxies" for the actual quantities of interest, without evidence on the quality of the proxies (McGlynn et al., 2008).

² Accounting for heterogeneity in the distribution of the inefficiency term is addressed below.

Newhouse's (1994) well-known skepticism about the "usefulness" of measuring the efficiency of health care services centered on the difficulty of measuring output in the health care sector. Newhouse (1994, p. 318) argues that, "The generic problem is the variation in quality of the product and its dimensionality; frontier techniques work best when the product is homogeneous and uni-dimensional, for example kilowatt-hours." Few efficiency studies specify "health" as the output; instead, most take the "medical care" approach. For hospital applications of SFA, the specified outputs include discharges and/or inpatient days (perhaps by payer type or level of care provided), surgeries, outpatient activities, etc. It is commonly argued that inpatient days or discharges are acceptable measures of output as long as adjustments are made for the complexity of cases (i.e., case-mix [which itself is based on diagnostic related groups—see the next paragraph]). However, the CMI used for US hospital studies is based on Medicare patients only, with the (implicit) assumption that this provides a uniform adjustment across all outputs for all sample observations. All-payer CMIs exist, but only for hospitals that use subscriber-based benchmarking services; thus these proprietary measures limit the "relative comparisons" that must be made to assess efficiency with SFA.

Aggregation bias is certainly a problem when applying SFA to health care. While this problem is not unique to health care research, it is likely exacerbated in this area given the complexity of health care "production." To address this, diagnostic related groups (DRGs) were developed to provide uniform measures of hospital activities that would facilitate comparisons across health care providers. DRGs group health care services on the basis of similar clinical characteristics (diagnoses) and resource utilizations (procedures). Widespread use of DRGs began with the implementation of Medicare's prospective payment system [PPS] in the U.S. in 1983, which mandated that payment for hospital services be based on DRGs. In an application of DEA to Danish hospitals, Olesen and Petersen (2002) measure cost efficiency based on 483 outputs (using DRGs). Making use of assurance regions, they argue that they cover the full output space without using fixed aggregation while controlling for heterogeneity in output. While ambitious, this still doesn't fully account for heterogeneity since a given DRG is not homogeneous with respect to resource use. In any case, this degree of disaggregation simply is not possible for parametric models like SFA.

Inputs are typically specified as various kinds of labor and capital. For hospital studies, capital is frequently proxied by the number of licensed and staffed beds. Production, however, is a process; thus, capital should be measured as a flow, not as a stock as in the case of number of beds. By including a stock measure of capital with flow measures of labor, the use of capital is overestimated and adjusted to both the level of capital (technical efficiency) and the optimal mix of capital and labor (allocative efficiency) will be biased. Furthermore, as more hospital activity is shifted from inpatient to outpatient care, the number of beds is not as relevant as it was in the past. As with outputs, measures of inputs also suffer from aggregation bias and presence of unobserved quality differences.

The AQA Alliance (2009) defines efficiency of health care as "a measure of cost of care associated with a specified level of quality of care." Failure to account for quality renders moot this definition of efficiency. Hussey et al. (2009) report that of 265 studies of health care efficiency that they reviewed, few controlled for quality. Without controlling for quality, what appears to be "efficient" production actually could be "low quality" production. While hospital efficiency studies routinely adjust for CMI, few adjust inputs or outputs for quality (there are exceptions, see, for example, Ferrier and Trivitt (2013)), nor is there agreement on how to adjust for quality of care.

Newhouse (1994) and others have noted that omitted variables bias is rife in health care efficiency studies. For example, most studies of US hospital efficiency do not include physicians as an input, since physicians have "admitting privileges" at hospitals but are not hospital employees (Veterans Affairs hospitals are an exception). Another omitted variable is

“quality.” Quality is multi-dimensional and difficult to quantify. Aggregation bias and proxy concerns are problematic for quality measures, even as more and more quality data become available. Furthermore, how to control for quality is an open question; e.g., should quality be included as a control variable or should inputs and/or outputs be adjusted for quality?

“Explaining” efficiency scores

In addition to measuring efficiency, many applications of SFA in health care examine the determinants or correlates of inefficiency. Doing so provides decision makers with insights on how efficiency might be improved. Rather than treating these analyses as two different steps, it is now recognized that for internal consistency, the measurement of efficiency and the analysis of its correlates must be done simultaneously. In addition to providing insight into the sources of inefficiency, this approach also offers another means for controlling for heterogeneity across health care providers. For example, the location of the truncated-normal distribution of the inefficiency component of the composite error term can be modeled as a function of internal or external factors thought to affect efficiency.³ Different values of the mean of the truncated-normal can have a big impact on the shape of the distribution.⁴

4. Conclusion

The SFA efficiency measures for the health care sector that appear in the academic literature are based largely on data that are relatively easy to obtain. Furthermore, the availability of software that allows the straightforward implementation of SFA models has increased. Thus, the use of SFA efficiency measures has become increasingly feasible. These considerations suggest that their use will grow, both in academic circles and among practitioners. However, feasibility and ease of efficiency measurement do not necessarily imply their use by decision makers. While frontier efficiency measures have been used in the incentive-based regulation of utilities, their use in the health care sector has remained limited (Lovell, 2006). Hussey et al. (2009, p. 784) conclude that this limited use is because, “Efficiency measures have been subjected to few rigorous evaluations of reliability and validity, and methods of accounting for quality of care in efficiency measurement are not well developed at this time. Use of these measures without greater understanding of these issues is likely to engender resistance from providers and could lead to unintended consequences.”

As Hollingsworth and Street (2006, p. 1055) observe, the “supply side” of applied research on health care efficiency “has grown into a thriving industry,” but the “demand side” is “less well-developed.” Hollingsworth and Street (2006) and McGlynn et al. (2008) appear to agree that the application of SFA to health care is relevant/important and timely/feasible, but the use of SFA by decision makers is inhibited by concerns about the reliability/scientific soundness of the approach and the lack of actionable results. The lack of reliability/soundness can be attributed to differences in the results produced by SFA and its main competitor, DEA; to the sensitivity of results to different formulations of the SFA model, including functional form, distributional assumptions, the choice of output and input measures; and the ability to account for heterogeneity across observations and separate it from inefficiency. The lack of “actionable” results stems in part from the fact that SFA is often applied at the organizational

³ The variance of the inefficiency term can also be modeled as a function of potentially influencing factors, offering a further means of accounting for heterogeneity.

⁴ Where to include control variables is an important modeling consideration as it affects identification. As noted here, they might be used to model the mean or the variance of inefficiency component of the error term or some combination of both (see Alvarez et al., 2006). Or as noted earlier, they might be included in the specification of the frontier function or in the correction for heteroscedasticity.

level rather than at the “production line” level—this “30,000 foot view” does not provide the specificity of insight needed to act.

Greater confidence in the usefulness of SFA can be gained by continuing to construct better models, improving measures of outputs and inputs to better capture the activities of health care providers while accounting for differences in “case-mix” and quality, performing analyses at more “micro,” and hence, actionable levels within health care organizations, greater use of confidence intervals rather than point estimates for reporting efficiency, and improvements in the identification of the forms and causes of inefficiency would all aid the diffusion of SFA from academia to industry and government. Working more closely with health care providers, payers, and regulators would help bring about these enhancements. Finally, some well-crafted case studies on how the use of SFA benefits health care providers could be great sales tools. The health care sector already uses a number of benchmarks against which performance can be assessed—DRGs, various financial ratios, proprietary measures of quality and safety (e.g., HealthGrades in the U.S). SFA is another tool that is available; it is now up to us to demonstrate the soundness and values of SFA efficiency measures.

Acknowledgements. I would like to thank both an anonymous referee and the Guest Editor Antonio Álvarez Pinilla whose constructive comments on an earlier draft of this paper improved it. I would also like to thank Vivian Valdmanis, from whom I have learned so much, for drawing me into this area of research. Any errors or omissions are my own.

References

- Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics, *Journal of Productivity Analysis*, 25(3), 201-212.
- AQA Alliance (2009) AQA principles of “efficiency” measures, available at <http://www.aqaalliance.org/files/PrinciplesofEfficiencyMeasurement.pdf> [Accessed on 24 April 2014]
- Chilingerian, J.A. and Sherman, H.D. (2011) *Health-care applications: from hospitals to physicians, from productive efficiency to quality frontiers*, in Cooper, W.W., Seiford, L.M., and Zhu, J. (Eds): *Handbook on Data Envelopment Analysis*, Springer, New York, 445-493.
- Dismuke, C.E. and Sena V. (1999) Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods, *Health Care Management Science*, 2(2), 107-116.
- Dor, A. (1994) Non-minimum cost functions and the stochastic frontier: on applications to health care providers, *Journal of Health Economics*, 13(3), 329–334.
- Ferrier, G.D. and Trivitt, J.S. (2013) Incorporating quality into the measurement of hospital efficiency: a double DEA approach, *Journal of Productivity Analysis*, 40(3), 337-355.
- Hollingsworth, B. (2003) Non-parametric and parametric applications measuring efficiency in health care, *Health Care Management Science*, 6(4), 203-218.
- Hollingsworth, B. (2008) The measurement of efficiency and productivity of health care delivery, *Health economics*, 17(10), 1107-1128.
- Hollingsworth, B., Dawson, P.J. and Maniadakis, N. (1999) Efficiency measurement of health care: a review of non-parametric methods and applications, *Health Care Management Science*, 2(3), 161-172.

- Hollingsworth, B. and Street, A. (2006) The market for efficiency analysis of health care organizations, *Health Economics*, 15(10), 1055-1059.
- Hussey, P. S., De Vries, H., Romley, J., Wang, M.C., Chen, S., Shekelle, P.G. and McGlynn, E.A. (2009) A systematic review of health care efficiency measures, *Health services research*, 44(3), 784-805.
- Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*, National Academy Press: Washington, D.C.
- Jacobs, R., Smith, P.C., and Street, A. (2006), *Measuring efficiency in health care: analytic techniques and health policy*, Cambridge University Press: Cambridge.
- Lovell, C.A.K. (2006) Frontier analysis in healthcare, *International Journal of Healthcare Technology and Management*, 7(1), 5-14.
- McGlynn, E.A., Shekelle, P.G., Chen, S., Goldman, D., Romley, J.A., Hussey, P.S., de Vries, H., Wang, M. et al. (2008) *Identifying, categorizing, and evaluating health care efficiency measures*, Agency for Healthcare Research and Quality: Rockville, MD.
- Newhouse, J.P. (1994) Frontier estimation: how useful a tool for health economics?, *Journal of Health Economics*, 13(3), 317-322.
- Olesen, O.B. and Petersen, N.C. (2002) The use of data envelopment analysis with probabilistic assurance regions for measuring hospital efficiency, *Journal of Productivity Analysis*, 17(1-2), 83-109.
- Rosko, M.D. and Mutter, R.L. (2008) Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness, *Medical Care Research and Review*, 65(2), 131-166.
- Sherman, H.D. (1984) Hospital efficiency measurement and evaluation: empirical test of a new technique, *Medical Care*, 22(10), 922-938.
- Triplett, J.E. (2013) *Health system productivity*, in Glied, S. and Smith, P.C. (Eds): *Oxford Handbook of Health Economics*, Oxford University Press: New York, 707-732.
- Wagstaff, A. (1989) Estimating efficiency in the hospital sector: a comparison of three statistical cost frontier models, *Applied Economics*, 21(5), 659-672.
- Wilson, G.W. and Jadlow, J.M. (1982) Competition, profit incentives, and technical efficiency in the provision of nuclear medicine services, *Bell Journal of Economics*, 13(2), 472-482.
- Worthington, A.C. (2004) Frontier efficiency measurement in health care: a review of empirical techniques and selected applications, *Medical Care Research and Review*, 61(2), 135-170.