Methods for Measuring IRS's Productivity

Kevin Daly and Jennifer Gravelle, U.S. Government Accountability Office

Many difficulties exist in measuring the productivity of public services. Public and private service providers often do not have an obvious measurable output. The output of a service, which can range from a haircut to heart surgery, usually involves a change in the condition of the consumer that is hard to quantify. IRS, by administering the tax code—which involves enforcing tax laws, collecting revenue, providing taxpayer services, etc.—provides a public service. Much research has been done on describing approaches to defining and measuring output of services in the public and private sector, as well as detailing the methods available for calculating productivity. In this paper, we apply these concepts to the issue of measuring productivity at IRS.

The organization of this paper is as follows. In the first section, we outline a framework, based on IRS's effects on the social cost of taxation, which can be used to aid in defining IRS's output. In the second section, we describe alternative methods for calculating productivity that are applicable to IRS. Lastly, we provide some limited illustrations of how these methods can be used to calculate productivity at IRS.

A Framework for Defining IRS Output

Because IRS produces a service, defining its output is difficult. The output of a service can be defined as a change in a good belonging to the consumer of the service or a change in the condition of the consumer. Housepainters provide a service by changing a good belonging to the consumer; doctors provide a service by changing the condition (health) of the consumer. (In either case, a change in the consumer's level of satisfaction occurs that can be viewed, for simplicity, as a change in the condition of the consumer.) Because services result in changes in the condition of the consumer, the unit of output is often intangible. In addition, because the transaction with the consumer that results in the service performed can include numerous interrelated services, the unit of output could be a complex bundle of services. This

interaction with the consumer also means that the output of the service is dependent on the consumer. For example, the output of a doctor depends on the original condition of the consumer and the consumer's actions during the service being performed. Similarly, the education output of a teacher is dependent both on characteristics of the consumers (students) and the actions the consumers take during the service.[1]

IRS provides a service by administering the tax code. IRS's service is intangible and complex because, like private sector services, it changes the condition of individuals (taxpayers) in a number of interrelated ways. Because IRS's effects are complex and intangible, it may be difficult to identify those effects that should be included in measures of output. A general description of the goal of IRS, which reinforces its mission statement (and is the view of the goal of IRS adopted in this paper), is to provide the service of administering the tax code at the lowest social cost. In pursuing this goal, IRS is constrained in what it has control over. IRS cannot change the tax code to increase revenue collections. However, there are elements of compliance and social cost of the tax system that IRS can affect.

While IRS administers the tax code, the tax rates and other provisions of the tax code are fixed and predetermined by Government policy. The tax code and characteristics of taxpayers determine the potential amount of revenue that can be raised. IRS's effect on revenue is limited to such factors as the effectiveness of its case selection for audits, the accuracy of the information it provides to taxpayers, the effect of its audit activity on voluntary compliance, etc. Therefore, unlike private firms where the objective is to maximize revenue net of private costs, IRS's goal, in a context where tax revenue represents societal transfers, is to maximize the level of compliance net of the social costs of administering the tax code.

The Social Costs of Administering the Tax Code

The social costs of administering the tax code are IRS's administrative costs, taxpayer's compliance burden, efficiency costs of tax avoidance and evasion, and perceptions of inequity.[2] Administrative costs are budgetary costs that arise directly from such IRS actions as processing returns and conducting audits. The resources used for IRS administration cost society the value of the output that could have been produced by these resources in alternative uses. Compliance burden is the cost that taxpayers incur complying with the tax code in terms of time and resources they use preparing returns and interacting with IRS. While IRS cannot change the code to reduce compliance burden, it can decrease taxpayer compliance burden by reducing the complexity of its forms and instructions and by providing better taxpayer services.

For example, overly complicated forms or poor instructions increase the time and resources a taxpayer must invest in order to comply with the tax system, while the implementation of e-filing can reduce the time a taxpayer takes to file his or her return.

Efficiency and equity costs, which are important elements of tax policy discussions concerning tax rates and bases, are also affected by IRS activities. The efficiency cost that IRS can affect is the social cost of taxpayers' decisions about the occupations that they follow, the investments that they make, and the resources that they use to avoid complying with the tax code. For example, a taxpayer may choose an occupation or investment based, at least in part, on the perception that IRS may be less likely to be able to collect tax liabilities incurred while engaged in these activities than others. To the extent that these decisions are affected by how taxpayers perceive that the tax code is being enforced by IRS, they represent distortions of economic decisionmaking and reduce the potential output of the economy. IRS may be able to affect this behavior (for example, through programs that induce voluntary compliance) and therefore affect efficiency costs associated with evasion or avoidance. The equity cost that IRS can affect is the social cost of taxpayers' perceptions of how fairly they are treated by IRS. Taxpayers may feel themselves less well off to the extent that they feel themselves subject to a tax system that is administered unfairly. IRS affects equity costs by how it selects taxpayers for enforcement activities and how accurately it applies the tax code to these taxpayers.

IRS's Objective Function

Formally, IRS's objective should be to maximize compliance net of the social cost of administration: Max C – S(A, B, E, I), where C denotes the level of compliance and S represents the social cost of administering the tax code.[3] The social cost, S, is a function of the administration costs, A; the compliance burden, B; the efficiency costs of tax avoidance, E; and taxpayers' perceptions of inequity, I.

Furthermore, IRS produces intermediate outputs that can affect compliance and the social costs: C = C(e, c, s); A = A(e, c, s); B = B(e, c, s); E = E(e, c, s); and I = I(e, c, s), where e is the level of enforcement produced by IRS, c is the collections of IRS, and s is the level of taxpayer service produced by IRS.

Formally, IRS objective function can be written as:

$$\text{Max } C(e, c, s) - S(\, A(e, c, s), B(e, c, s), E(e, c, s), I(e, c, s)\,)$$

which has the following first-order conditions for optimization:

$$\partial C/\partial e = \partial S/\partial A * \partial A/\partial e + \partial S/\partial B * \partial B/\partial e + \partial S/\partial E * \partial E/\partial e + \partial S/\partial l * \partial l/\partial e$$

$$\partial C/\partial c = \partial S/\partial A * \partial A/\partial c + \partial S/\partial B * \partial B/\partial c + \partial S/\partial E * \partial E/\partial c + \partial S/\partial l * \partial l/\partial c$$

$$\partial C/\partial s = \partial S/\partial A * \partial A/\partial s + \partial S/\partial B * \partial B/\partial s + \partial S/\partial E * \partial E/\partial s + \partial S/\partial l * \partial l/\partial s$$

Thus, IRS should produce enforcement until the marginal increase of enforcement on compliance equals the marginal increase of enforcement on social costs and similarly for taxpayer service and collections.

The objective function emphasizes the need to balance compliance gains and social costs. Achieving this balance in practice is difficult because of the complicated nature of the service that IRS provides. Enforcement, taxpayer service, and collection processes are themselves functions of a variety of activities that IRS performs and therefore can be increased or decreased by the mix of these activities. For example, enforcement is a function of various types of audits, criminal investigations, and appeals as follows:

$$e = e(a_1, a_2, \ldots, a_n; c, ap)$$

where a_i denotes specific type of audit, c, criminal investigations, and ap, appeals. Furthermore, as the objective function indicates, compliance is a function of more than just enforcement activities such as audits; it is also a function of taxpayer service and collection. Services are defined by the following function:

$$s = s(t_1, t_2, \ldots, t_m)$$

where t_i denotes a specific type of taxpayer service action. For example, phone service, publications, walk-in site assistance are all forms of taxpayer assistance. In this case, t_i might equal phone calls answered or taxpayer questions answered correctly. Collections are a function of paper returns, p, and electronic returns processed, r, and other collection activities such as levies and seizures, l, as follows:

$$c = c(p, r, l)$$

At this level of specific activities, the mix should be done for internal efficiency and adjusted for the quality and complexity of the action. Lastly, each of these actions, a_i, t_i, etc., are functions of capital and labor inputs.

Measuring Outputs and Inputs

There are two general approaches for defining output in service industries that address the problem of how to measure outputs: the transactions and outcomes approach. Transactions are the procedures, activities, or outputs that produce an outcome. An outcome is the final result or consequence of the service performed.

If output is defined in terms of its effects on compliance and social cost (the outcomes approach), IRS needs to identify, measure, and specify tradeoffs among these effects. On the other hand, if output is defined in terms of the workload of IRS (the transactions approach), output measurement may be easier, but outputs are less directly linked to their effects on taxpayers and the goals of the agency.[4]

The transactions approach centers on defining measures that reflect the work done, rather than the consequences of that work. The transactions approach produces internal or operational measures, concerned chiefly with the technical efficiency of the organization. An example of a transactions-based productivity measure in IRS would be cases closed per full-time equivalent employees (FTE's). Applying this method correctly entails adjusting for the quality and complexity of the transaction. For example, an increase in cases closed per FTE would not indicate an increase in productivity if the increase occurred because FTE's were shifted to less complex cases or the examiner allowed the quality of the case review to decline in order to close cases more quickly.[5]

The outcomes approach centers on defining measures that reflect the results from the service performed. The results of the service are the effects on the consumer, or, in the case of IRS, on the taxpayers who receive the service. The outcomes approach is preferred because it focuses, not merely on the internal efficiency of the organization, but on the organization's impact on the people it serves. (See Table 1 for example of outputs under the transactions and outcomes approach.) However, outcomes are difficult to measure.[6] As stated above, services generally affect the people who receive them in complex, interrelated, and intangible ways. For example, IRS's audit activities may impose costs on the taxpayers being audited but may also affect other taxpayers by increasing involuntary compliance or their perceptions of the fairness of the tax system. An outcomes-based measure of IRS's output of services would capture these and other effects on taxpayers (described below). In addition, identifying the portion of the outcome due to consumer effects may be difficult. For example, the time it takes IRS to complete an

exam depends on the complexity of the return, which will differ depending on the taxpayer.

Table 1: Examples of Outputs of Public Sector Services Using the Transactions and Outcomes Approach

Service	Outputs	
(Purpose)	Transactions	Outcomes
Corrections (House/rehabilitate offenders)	Clothe inmates Serve meals Patrol cell blocks	Reduce crime Protect society
Education (Educate students)	Conduct classes Give tests Serve meals Operate school buses	Increase literacy, human capital
Fire (Put out and prevent fires)	Maintain fire trucks Train firefighters	Reduce fire losses and deaths
IRS (Collect taxes)	Produce and distribute tax forms Process Returns Answer calls Perform exams	Increase compliance and equity Decrease compliance burden and efficiency costs

Source: Adapted from BLS (1998)

In order to implement an outcomes approach, IRS can try to measure outcomes directly or use transactions that are proxies for outcomes. For example, IRS could try to estimate the effect of its enforcement activities on voluntary compliance or it could use transactions like audit rates, which may be correlated with voluntary compliance. These transactions would serve as proxies for outcomes that cannot be directly measured. In either case, multiple indicators of output would be necessary to capture the full range of effects that IRS has on taxpayers.

Measuring Productivity

Since productivity is the efficiency with which inputs are used to produce outputs, measuring productivity is difficult for services because defining and measuring output are difficult. Depending on the type of output measures used, different types of methods for calculating productivity and its changes over time may be required.

While there may certainly be a number of cases where a single output to input ratio provides accurate information for many outputs, a single ratio index does not capture all the complexities and changes in the service produced over time. In addition, single output to input ratios cannot provide more

comprehensive productivity measures that cover a range of different outputs and multiple inputs. While there are a number of methods for combining multiple outputs and inputs, including different types of weighted indexes and stochastic frontier analysis which uses regression methods to estimate costs functions, this paper focuses on the method of Data Envelopment Analysis (DEA).[7]

DEA, which has been gaining popularity in the field of productivity measurement, has been used in a wide range of applications from measuring efficiency in the bank industry and hospitals to use in governments to measure the efficiency of certain programs.[8]

DEA is a nonparametric estimation technique that uses a linear program to estimate a production function from the most efficient producing units, referred to as decision-making units (DMU's) in the literature. It then assigns efficiency scores to the remaining producing units according to how far they are from the estimated efficient frontier. Formally, the output-oriented linear program model for each DMU is:

$$(D^t(x_i^t, y_i^t))^{-1} = \max \theta$$

subject to

$$\sum_{s=1}^{S} l_s y_{ms} \geq q y_{mi} \qquad m = 1, 2, \ldots, M;$$

$$\sum_{s=1}^{S} l_s x_{ns} \leq x_{ni} \qquad n = 1, 2, \ldots, N;$$

$$\sum_{s=1}^{S} l_s = 1$$

$$\lambda_s \geq 0 \qquad s = 1, 2, \ldots, S;$$

where y_{mi} and x_{ni} are the mth and n^{th} output and input used by DMU_i.

In the output-oriented method, the solution $\theta*$ is the scalar that expands output as far as possible such that that output is still producible with the fixed level of inputs x. If $\theta*$ is =1, then the DMU is considered to be efficient because output could not be expanded any more without increasing the level of inputs. A solution value of $\theta* > 1$ indicates an inefficient DMU, relative to the efficient DMU's, since more output could currently have been produced with the same level of inputs.

The inverse of this scalar value is equal to distance function, $D^t(x_i^t, y_i^t)$. Thus, when $\theta^* > 1$, $D^t(x_i^t, y_i^t) < 1$ indicating inefficiency. The distance functions are a measure of how far an output and input combination are from a production frontier. The use of distance functions is particularly important in measuring productivity in public services where prices are not available.[9] The Malmquist index is one such measure of productivity that does not rely on prices but rather on changes in the distance functions over time which, as can be seen from the linear program, rely only measures of outputs and inputs.[10] The Malmquist index measuring productivity change over a given time period is the geometric mean of the ratio of distance functions in each period:

$$\left(\frac{D_i^o(x_i^t, y_i^t)}{D_i^o(x_i^o, y_i^o)} \cdot \frac{D_i^t(x_i^t, y_i^t)}{D_i^t(x_i^o, y_i^o)} \right)^{1/2}$$

where 0 denotes the current period and t the future period. Thus, the Malmquist index produces a measure of productivity change over time, either period to period or relative to a base year. The Malmquist index can be further decomposed into efficiency change and technology change:

$$\left(\frac{D_i^t(x_i^t, y_i^t)}{D_i^o(x_i^o, y_i^o)} \right) \left[\frac{D_i^o(x_i^t, y_i^t)}{D_i^t(x_i^t, y_i^t)} \cdot \frac{D_i^o(x_i^o, y_i^o)}{D_i^t(x_i^o, y_i^o)} \right]^{1/2} = E^t \cdot T^t$$

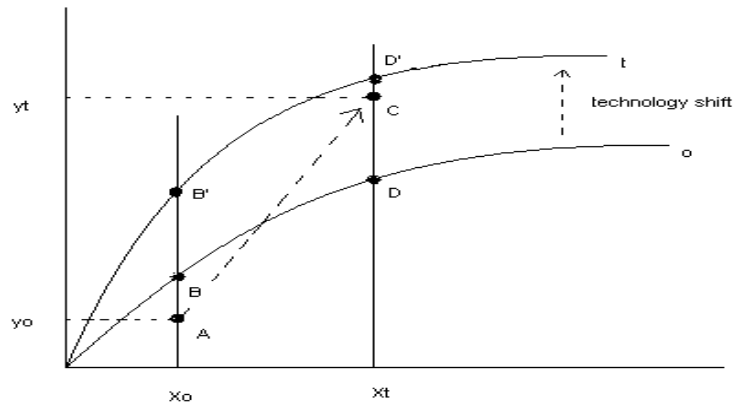
The decomposition allows productivity changes to be measured in terms of the efficiency change, holding technology constant, and the effect of technology on the ability to produce. As Figure 1 shows, the technology term reflects only the change in technology as measured by

$$\left(\frac{Cx_t / Dx_t}{Cx_t / D'x_t} \cdot \frac{Ax_o / Bx_o}{Ax_o / B'x_o} \right)^{1/2} = \left(\frac{D'x_t}{Dx_t} \cdot \frac{B'x_o}{Bx_o} \right)^{1/2}$$

, the geometric mean of technology shifts in the two time periods. (In the figure, $D'x_t/Dx_t$ represents the distance from D to D', and $B'x_0 / Bx_0$ represents the distance from B to B'.) This measure is independent of the efficiency of the firm in either period.

The efficiency term change is represented in figure 1 by $\left(\frac{Cx_t / D'x_t}{Ax_o / Bx_o} \right)$, the change in efficiency relative to current production technology. (In the figure, $Cx_t/D'x_t$ represents the distance from C to D', and Ax_0 / Bx_0 represents the distance from A to B.) The final change in productivity is shown by the move from A to C, which included changes in efficiency and a change in the production technology.

Figure 1: Decomposition of Malmquist Index



DEA's ability to include a variety of different outputs including qualitative outputs makes broad productivity measurement more feasible. Its nonparametric methods eliminate the need to specify functional forms, as in the case of stochastic frontier analysis. However, unlike stochastic frontier analysis, DEA can confound inefficiency with stochastic noise and does not have readily available measures of statistical significance and goodness of fit.[11] DEA's method of optimally assigning weights is thought to provide a "best case" scenario of efficiency. It constructs the frontier by allowing the DMU's, in effect, to choose the weights for their various outputs that make them appear most productive. Since all DMU's are free to choose these weights, the frontier is defined by the best practice DMU's in the organization or the industry. However, this procedure can make DEA sensitive to outliers. It should also be understood that DEA estimates the efficient production frontier according to the observations provided, and therefore, in an absolute sense, all the DMU's could be inefficient.

Illustration of DEA and Malmquist Indexes at IRS

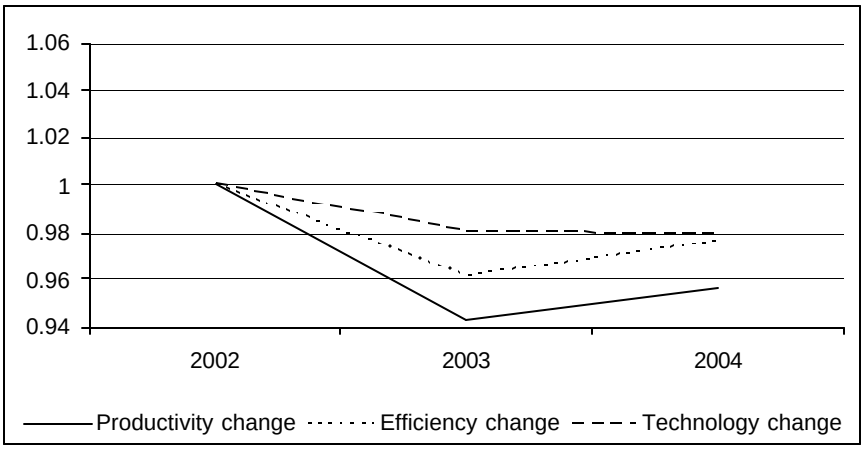
DEA is most useful when there are a large number of DMU's. While the illustrations we provide are based on data that include a fairly small number of DMU's, IRS may in time be able to effectively use DEA in measuring productivity of outputs that are similar across divisional units. In addition, data available at area levels in specific divisions, for example Wage and Investment, may provide a number of DMU's along with a variety of outputs, from compliance to taxpayer service, which could be used to obtain broad estimates of

productivity across different types of functions.

Our illustrations use exam and FTE data from the 5 industry groups within the Large and Mid-Sized Business (LMSB) division. The exam outputs are broken into 5 types of exams: corporate exams under $10 million, individual exams under $100,000, individual exams over $100,000, business industry exams, and coordinated industry exams. It should be noted that, ideally, a larger variety of outputs over a longer period would be preferable. These measures are therefore only illustrations, and we do not intend these to be definitive measures of productivity at IRS. We also do not explore in depth the productivity changes presented in these illustrations as they are used only to provide examples of what type of information DEA and Malmquist indexes could provide IRS.

Figure 2 shows productivity change over the period 2002 to 2004. As can be seen, the use of DEA to estimate Malmquist indexes allows productivity change to be broken into changes in efficiency and changes in technology. The illustration in Figure 2 suggests that, while technology declined slightly over the period, larger declines in efficiency accounted for much of the change in productivity. One of the benefits of using the Malmquist index is the ability to separate out changes in technology, which may easily come from factors beyond IRS's control. For example, changes in the rules and regulations that require more work for a given audit could be represented by a shift of the production function inward, indicating that, with the same amount of inputs, fewer outputs could be produced.

Figure 2: Malmquist Index with All outputs



As mentioned earlier, DEA's method of optimally assigning weights is thought to provide a best case scenario of efficiency. However, the free movement of weights could mask other changes. DEA assigns the most weight to those outputs for which the DMU compares favorably and the least weight on those outputs it does not efficiently produce. This optimal assignment, while representing a benchmark for efficiency, may not reflect the preferences of the organization. The organization may believe that some outputs are more important than others and should therefore have a greater weight. Weight restrictions can reflect these preferences, and the weights can also be varied simply for the purpose of analyzing the sources of productivity change.

Figure 3 shows the same outputs with an addition restriction that the total weights assigned to the business industry and coordinated industry exams be larger than the total weights assigned to low-income individual, high-income individual, and low-asset corporate exams. As can be seen, while technology change was little affected by the weight restriction, efficiency change differed dramatically. With the weight restriction, efficiency change over the period is largely positive so that total productivity change over the period is positive.

In general, LMSB was able, over this period, to be more productive in individual and corporate examinations—to do more exams per FTE—than in business industry and coordinated industry exams. Individual and corporate exams set a benchmark not matched by business industry and coordinated exams which caused the overall decline in productivity. The heavy weights that DEA may have placed on the individual and corporate exams seems to have masked large increases over this period in the number of business and coordinated industry exams performed.

Figure 3: Malmquist Index with All Outputs and a Weight Restriction

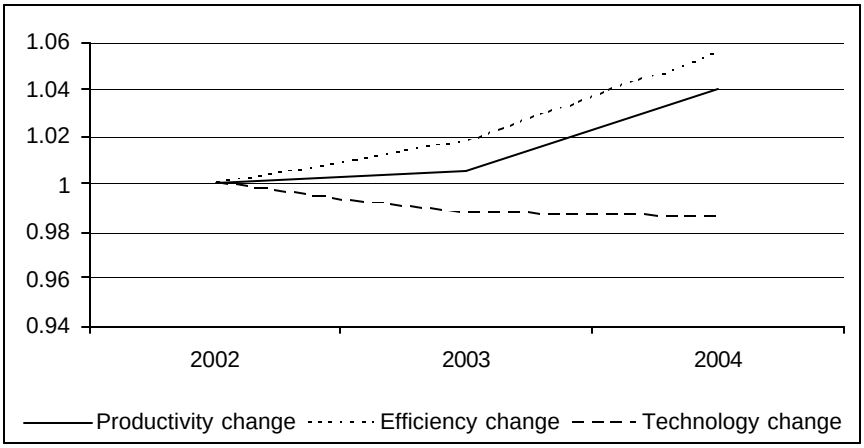


Figure 4 shows productivity, efficiency, and technology change over the period 2002 to 2004 for only the business industry and coordinated industry exams. As can be seen, even if technology is decreasing (or remaining close to one), large increases in efficiency can override decreasing technology and produce increases in productivity.

Figure 4: Malmquist Index with Only Business Industry and Coordinated Industry Exams

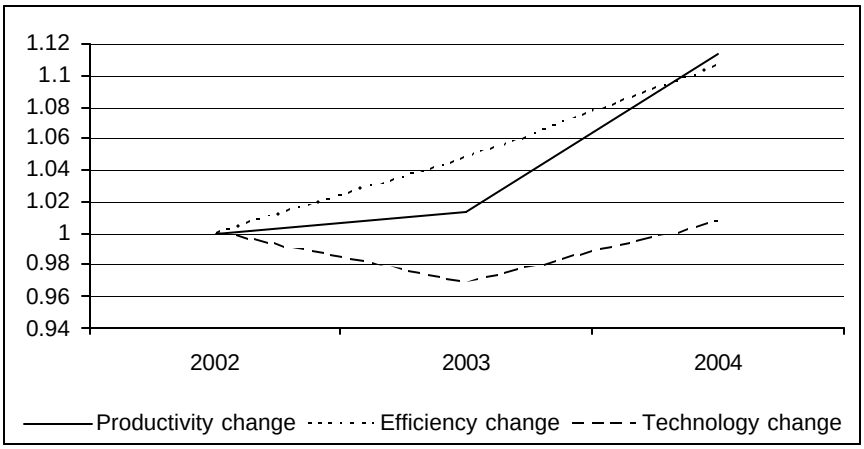
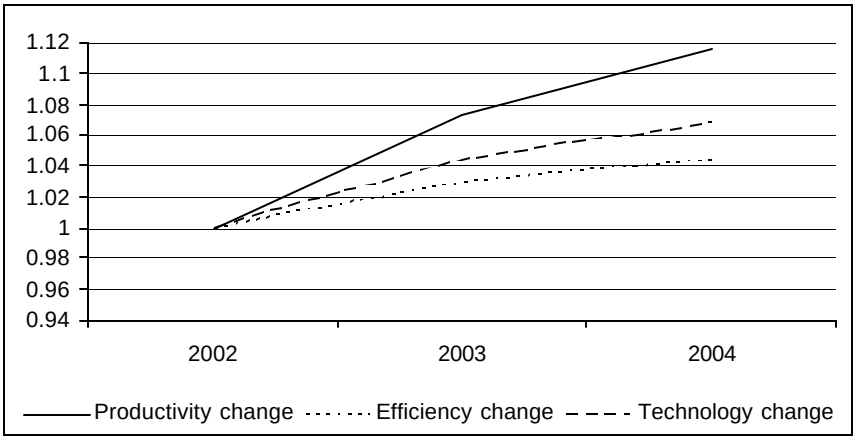


Figure 5 shows the effect of including quality scores. The inclusion of quality scores shows technology now increasing over the period, as both business industry and coordinated industry quality scores generally increased. In the prior examples that included only exams of different types as outputs, the decline in technology represented a downward shift in the frontier. The best practice DMU's appear less efficient in terms of number of exams closed per FTE. However, when the exams are adjusted for quality, by adding a quality score as a separate qualitative output in the analysis, the shift of the frontier is more than offset.

Figure 5: Malmquist Index Only Business Industry and Coordinated Industry Exams and Quality Scores



Conclusion

Measuring productivity in services is difficult because defining the output of the service is difficult. IRS provides the service of administering the tax code. The service it provides is the greatest level of compliance at the least social costs. Compliance and social costs are aspects of IRS's service that define its output.

To obtain overall measures of productivity at IRS, a variety of outputs should be used to capture the different types of functions performed in administering the tax code. While there are a number of ways of combining multiple outputs, Data Envelopment Analysis has been gaining popularity in productivity measurement. DEA allows productivity changes to be decomposed into changes in efficiency and changes in technology. The ability to separate changes in productivity into its components could provide important information about the causes of productivity change. The ability to restrict weights allows deeper exploration into the causes of productivity changes and could, therefore, ultimately provide more information about how to counter decreasing productivity or continue increases in productivity. In the end, the benefit to IRS of using this and other methods of measuring productivity is to provide increased information on which to base decisions that affect how IRS operations are performed.

Endnotes

[1] See Sherwood (1994) for a discussion of the requirements for, and the difficulties of, measuring the output of services in the private sector and the Bureau of Labor Statistics (1998) for measurement requirements and difficulties in the public sector.

[2] See Slemrod and Yitzhaki (1996) for a description of the social costs of administering the tax laws. They describe these costs in the context of the marginal cost of funds (MCF) approach to evaluating changes in tax law and tax administration. This MCF approach separates the ultimate benefits of spending funded by the taxes from the costs of collecting the taxes. Specifically, in their model, social welfare is maximized (or social costs minimized) using tax and administrative instruments subject to the constraint that the tax agency raises a given amount of revenue. The solution to their optimization problem describes the social costs of marginal variations of the instruments—the MCF's of the instruments—which can be used to identify welfare-improving tax and administrative changes. (This MCF approach was initially applied to tax changes by Ahmad and Stern (1984) and expanded to include tax administration changes by Slemrod and Yitzhaki (1987) and Mayshar (1991). Recently, Slemrod and Yitzhaki (2001) have also argued for including a measure of MCF in cost-benefit evaluations of individual spending projects.) The few studies that provide empirical estimates of the social costs of tax collection deal only with compliance costs—for example, see Slemrod and Sorum (1984) and Blumenthal and Slemrod (1992)—and these studies do not link changes in compliance costs to specific tax agency activities.

[3] This objective function for IRS is consistent with a variant of the standard Ramsey model. IRS's objective function can be derived from a more general model of maximizing a social welfare function over tax policy instruments such as tax rates and bases (considered fixed) and tax administration instruments (considered variable) subject to a revenue constraint. It is therefore also consistent with the Slemrod and Yitzhaki MCF approach described in endnote 2.

[4] Both the Bureau of Labor Statistics (1998) and Fisk and Forte (1997) describe the difficulty of measuring outcomes and note that the Federal productivity measurement program used multiple indicators of final outputs (similar to transactions) rather than outcomes in its productivity measures. However, as Nyhan and Martin (1999) report, recent initiatives, especially the Government Performance and Results Act (GPRA),

have led to increased emphasis on effectiveness (outcome) performance measurement.

[5] For a discussion of this issue, see GAO (2004).

[6] In the case where outcomes can be measured directly, there are three main advantages to measuring outcomes: 1) only the outcome need be measured as opposed to measuring all the elements that go into producing the outcome; 2) outcomes can account for changes in quality reflected in an increased probability of a given outcome; and 3) outcomes can also include product innovations.

[7] DEA was introduced by Charnes et al. (1978). Their method, which assumed a constant returns to scale technology, was later modified by Banker et al. (1984) and Chavas and Cox (1990) to include variable returns to scale technologies.

[8] DEA has been applied extensively in both the public and private sector. See Seiford (1990) for a survey of the DEA literature.

[9] For examples of the literature using distance functions as measures of relative efficiency, see Valdmanis (1992) and Ruggiero and Vataliano (1999).

[10] See Caves et al. (1982) and Sudit (1995) for descriptions of the Malmquist index, its history, and its relationship to other indexes. Fare et al. (1985) first directly estimated the Malmquist productivity change index as a ratio of distance functions. They also introduced—see Fare et al. (1994)—the decomposition of the index into technical efficiency change and technological change. For examples of the literature using ratios of distance functions as measures of productivity change, see Wheelock and Wilson (1999) and Bjurek and Hjalmarsson (1995).

[11] For an example of the stochastic frontier approach, see Berger and Mester (1997) who estimate a cost frontier for the U. S. banking industry and use it to analyze productivity change and its decompositions. For an assessment of the advantages and disadvantages of parametric and nonparametric approaches to frontier analysis, see Seiford and Thrall (1990). Ruggiero and Vitaliano (1999) compare the results of an analysis of public school efficiency using DEA and a stochastic cost frontier. See Grosskopf (1996) for a review of methods for statistical inference used with DEA. Linna (2000) applies bootstrapping techniques to develop confidence intervals for a Malmquist productivity change index and its decompositions in his study of productivity change in Finnish hospitals.

References

Ahmad, E., and Stern, N., "The Theory of Reform and Indian Indirect Taxes" *Journal of Public Economics*, December 1984, 25(3), pp. 259-298.

Banker R.; Charnes, A.; and Cooper, W. (1984), "Models for the Estimation of Technical and Scale Efficiencies in Data Envelopment Analysis," *Management Science*, 30, pp. 1078-1092.

Berger, A. and Mester, L. (1997), "Efficiency and Productivity Change in the U.S. Commercial Banking Industry: A Comparison of the 1980's and 1990's," *Federal Reserve Bank of Philadelphia Working Papers*, #97-5.

Blumenthal, M. and Slemrod, J., "The Compliance Cost of the U.S. Individual Income Tax System : A Second Look After Tax Reform," *National Tax Journal*, June 1992, 45, pp. 185-192.

Bjurek, H. and Hjalmarsson, L. (1995), "Productivity in Multiple Output Public Service: A Quadratic Frontier Function and Malmquist Index Approach," *Journal of Public Economics*, 56, pp. 447-460.

Caves, D.; Christensen, L.; and Diewert, W. (1982), "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity," *Econometrica*, 5, pp. 1393-1414.

Charnes, A; Cooper, W.; and Rhodes, E. (1978), "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2, pp. 429-444.

Chavas, J. and Cox, T.. "A Non-Parametric Analysis of Productivity: The Case of U.S. and Japanese Manufacturing," *The American Economic Review*, June 1990, 80(3), pp. 450-464.

Fare, R.; Grosskopf, S.; and Lovell, C.A.K., *The Measurement of Efficiency of Production.* Boston: Kluwer-Nijhoff Publishing, 1985.

Fare, R.; Grosskopf, S.; Norris, M.; and Zhang, Z., "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *The American Economic Review*, March 1994, 84(1), pp. 66-83.

Fare, R.; Grosskopf, S.; and Norris, M., "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Reply," *The American Economic Review*, December 1997, 87(5), pp. 1040-1043.

Fisk, D. and Forte, D., "The Federal Productivity Measurement Program: Final Results," *Monthly Labor Review*, May 1997, 120(5), pp. 19-28.

Grosskopf, S. (1996), "Statistical Inference and Nonparametric Efficiency: A Selective Survey," *Journal of Productivity Analysis*, 7, pp. 161-176.

Linna, M., "Health Care Financing Reform and Productivity Change in Finnish Hospitals," *Journal of Health Care Finance*, Spring 2000, 26(3), pp. 83-100.

Mayshar, J. (1991), "Taxation with Costly Administration," *Scandinavian Journal of Economics*, 93(1), pp. 75-88.

Nyhan, R. and Martin, L., "Comparative Performance Measurement: A Primer on Data Envelopment Analysis," *Public Productivity & Management Review*, March 1999, 22(3), pp. 348-364.

Ray, S. and Desli, E., "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment," *The American Economic Review*, December 1997, 87(5), pp. 1033-1039.

Ruggiero, J. and Vitaliano, D., "Assesssing the Efficiency of Public Schools Using Data Envelopment Analysis and Frontier Regression," *Contemporary Economic Policy*, July 1999, 17(3), pp. 321-331.

Seiford, L. (1996), "Data Envelopment Analysis: An Evaluation of the State of the Art, 1978-1995," *Journal of Productivity Analysis*, 7, pp. 99-137.

Seiford, L. and Thrall, R. (1990), "Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis," *Journal of Econometrics*, 46(1/2), pp. 8-38.

Sherwood, M., "Difficulties in the Measurement of Service Outputs," *Monthly Labor Review*, March 1994, pp. 11-19.

Slemrod, J. and Sorum, N., "The Compliance Cost of the U.S. Individual Income Tax System," *National Tax Journal*, December 1984, 37(4),

pp. 461-474. Also in NBER Working Paper #1401, July 1984.

Slemrod, J. and Yitzhaki, S. (1987), "Optimal Size of a Tax Collection Agency," Scandinavian Journal of Economics, 89(2), pp. 183-92. Also in NBER Working Paper #1759.

Slemrod, J. and Yitzhaki, S., "The Costs of Taxation and the Marginal Efficiency Costs of Funds," *IMF Staff Papers*, March 1996, 43(1), pp. 172-198.

Slemrod, J. and Yitzhaki, S., "Integrating Expenditure and Tax Decisions: The Marginal Cost of Funds and The Marginal Benefit of Projects," *National Tax Journal*, June 2001, 54(2), pp.189-201. Also in NBER Working Paper #8196, March 2001.

Sola, M. and Prior, D., "Measuring Productivity and Quality Changes Using Data Envelopment Analysis: An Application to Catalan Hospitals," *Financial Accounting & Management*, August 2001, 17(3), pp. 219-245.

Sudit, E. (1995), "Productivity Measurement in Industrial Operations," *European Journal of Operational Research*, 85, pp. 435-453.

U.S. Bureau of Labor Statistics, "Measuring State and Local Government Labor Productivity: Examples from Eleven Industries," Bulletin 2495, June 1998.

U.S. Government Accountability Office, "Tax Administration: Planning for IRS's Enforcement Process Changes Included Many Key Steps But Can Be Improved," GAO-04-287, January 20, 2004, Washington, D.C.

Wheelock, D. and Wilson, P., "Technical Progress, Inefficiency, and Productivity Change in U.S Banking, 1984-1993," *Journal of Money, Credit, and Banking*, May 1999, 31(2), pp. 212-234.

Valdmanis, V., "Sensitivity Analysis for DEA Models: An Empirical Example Using Public vs. NFP Hospitals," *Journal of Public Economics*, July 1992, 48(2), pp 185:205.

Worthington, A. and Dollery, B., "Measuring Efficiency in Local Governments' Planning and Regulatory Function," *Public Productivity and Management Review*, June 2000, 23(4), pp. 469-485.