Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP12/2010

Sensitivity analysis of network DEA illustrated in branch banking

Necmi K Avkiran

Date: December 2010

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Sensitivity analysis of network DEA illustrated in branch banking

Necmi K Avkiran *

* Send correspondence to Associate Professor Necmi K Avkiran, UQ Business School, The University of Queensland, Brisbane QLD4072, Australia (e-mail: n.avkiran@uq.edu.au), tel: +(61 7) 334 63282; fax: +(61 7) 334 68166.

Sensitivity analysis of network DEA illustrated in branch banking

Abstract

Users of data envelopment analysis (DEA) often presume efficiency estimates to be robust. While traditional DEA has been exposed to various sensitivity studies, network DEA (NDEA) has so far escaped similar scrutiny. Thus, there is a need to investigate the sensitivity of NDEA, further compounded by the recent attention it has been receiving in literature. NDEA captures the underlying performance information found in a firm's interacting divisions or sub-processes that would otherwise remain unknown. Furthermore, network efficiency estimates that account for divisional interactions are more representative of a dynamic business. Following various data perturbations overall findings indicate positive and significant rank correlations when new results are compared against baseline results - suggesting resilience. Key findings show that, (a) as in traditional DEA, greater sample size brings greater discrimination, (b) removing a relevant input improves discrimination, (c) introducing an extraneous input leads to a moderate loss of discrimination, (d) simultaneously adjusting data in opposite directions for inefficient versus efficient branches shows a mostly stable NDEA, (e) swapping divisional weights produces a substantial drop in discrimination, (f) stacking perturbations has the greatest impact on efficiency estimates with substantial loss of discrimination, and (g) layering suggests that the *core inefficient cohort* is resilient against omission of benchmark branches. Various managerial implications that follow from empirical findings are discussed in conclusions.

Key words: Network DEA; sensitivity analysis; robustness; banking

1. Introduction

Network data envelopment analysis (NDEA) has recently been enjoying renewed academic interest but robustness of estimates remains untested. NDEA is a nonparametric, sophisticated approach to relative performance modeling that takes advantage of interrelated production frontiers (see Färe and Grosskopf [1] and [2]). Network DEA captures the underlying performance information found in a firm's interacting divisions or sub-processes that would otherwise remain unknown to management. For example, NDEA could help identify potential merger synergies trapped in organizational sub-units that are often monitored at a managerial accounting level but normally not investigated as interacting components of a larger network. Network efficiency estimates produced by NDEA that account for divisional interactions are more representative of a dynamic business than static measures that report overall performance without opening the *black box* of production. In short, NDEA is a useful technique to keep in the toolkit for organizational performance measurement because there is a need for using sophisticated efficiency measures that can accommodate the complexity of today's organizations.

Having established the case for NDEA, we now turn to literature on sensitivity of traditional DEA efficiency estimates to data perturbations. According to Chapparo *et al.* [3, p.637], "… advocates of DEA are often implicitly suggesting the pre-eminence of certain criteria - in particular robustness…". Examples of papers that have attempted to systematically show the key implications for efficiency estimates of various data perturbations include Smith [4], Ruggiero [5], Chapparo *et al.* [3], Seiford and Zhu [6], Galagedera and Silvapulle [7], and Avkiran [8]. Sensitivity analyses often investigate the stability or robustness of results to *changes* in the sample size, number of variables in the analysis, importance of inputs, and correlation between inputs. Findings indicate that DEA estimates of efficiency are closest to the so-called 'true' levels of efficiency estimated using a Cobb-Douglas type function when production process is simple in terms of the variables used and sample sizes are large in relation to the number of variables. However, in network DEA, the production process cannot be modeled as a small number of inputs leading to a single output. This is because

exogenous inputs entering the black box are often apportioned to various divisions, linked by intermediate products, all of which lead to final outputs exiting the black box (see Figure 4 in Färe and Grosskopf [2], and Figure 2 in Avkiran [9]).

The choice of variables in DEA is often driven by managerial focus (expert opinion), experience, production theory, or some combination of these approaches. Issues often overlooked in literature reporting applications of DEA include the technique's exploratory nature, sensitivity of efficiency estimates to the composition of the efficient frontier, and an efficient frontier which, in turn, can be sensitive to the choice of variables. Thus, given the many opportunities for misinterpreting DEA results, studies that undertake a systematic critical investigation of models that are becoming more popular, such as the network slacks-based measure (NSBM) by Tone and Tsutsui [10], are worthwhile.

The current paper is the first attempt to systematically test the robustness of estimates generated by weighted NSBM to various data perturbations and openly discuss various challenges awaiting attention of researchers. The motivation is to bring more confidence to findings based on network DEA, and hopefully, a wider adoption of this technique in organizational decision-making. Section 2 details the conceptual framework behind branch production and NSBM. Section 3 introduces the methodology used in generating data, discusses various types of perturbations and outlines jackknifing. Results are reported in section 4, followed by a discussion in section 5. The paper concludes in section 6 with a summary of key findings and a discussion of managerial implications.

2. Conceptual framework

2.1 Bank branch production

We want to illustrate sensitivity analysis of network DEA in a business setting where a homogeneous organizational network structure can be easily identified. We fulfill this objective by conceptualizing an intuitive and basic framework describing the main production processes in a retail bank branch. We begin by showing in Figure 1 how traditional DEA would treat a branch as a black

box. That is, a set of exogenous inputs (resources) enter a branch, only to emerge as a set of final outputs. Thus, traditional DEA does not explicitly identify any of the key sub-processes engaged by divisions found within a branch.

[Insert Figure 1 about here]

In conceptualizing a network structure, let us now imagine what is likely to happen when a customer walks into a branch. For example, some banks employ a numbered service counter system. The customer first picks up a ticket and waits to be called to one of the teller windows. Once at the window, the customer may initiate one of many simple transactions such as depositing or withdrawing money, opening a term deposit account, purchasing some foreign currency, etc. (see Figure 2). Higher level transactions such as an inquiry about which of the many types of credit cards would be suitable for the customer, taking out a home loan, starting a tax-effective retirement savings account, etc., would normally be referred to one of the relationship bankers who work in a separate room. This key production link between two functional divisions (i.e., tellers and relationship bankers) is known in network DEA literature as an *intermediate product* in acknowledgement of a transaction that starts as an output from one division and becomes an input into another division. The successful sale of products and services that require a more in-depth interview with a customer depends on the resources allocated to the division of relationship bankers, including interpersonal skills acquired in training that is brought to the encounter. Of course, one cannot expect all the referrals from the *division of tellers* to be turned into sales in the *division of relationship bankers*.

[Insert Figure 2 about here]

As part of the exercise of conceptualizing realistic branch production processes, we need to make some basic assumptions about the relationships between the variables before generating data. For example, according to Smith [4, p.236], “In most practical applications of DEA, inputs are highly correlated because they are all related to the scale of operations of the DMU.” Assuming a performance measurement period of one year, the arguments developed below detail the relationships

between the modeled variables and divisions originally outlined in Figure 2 (the mathematical algorithm is shown in section 3.2).

- Various assumed data ranges, while arbitrary for convenience, follow reasonable expectations for bank branch operations.

- Average flow of customer traffic is assumed to manifest itself in *labor* input measured in full-time equivalent or FTE. As we shall see below, labor becomes the key variable driving the rest of the variables. Total labor for each branch is randomly determined in the range of 4 FTE (small branch) to 16 FTE (large branch).

- We assume the division of tellers uses more labor and set this relationship to randomly vary between 120 - 150%. For example, if we focus on a 10 FTE branch and assume tellers use 130% more labor, then division-level labor emerges as 5.65 FTE (tellers) and 4.35 FTE (relationship bankers).

- Values for *labor* and *capital* within divisions are positively correlated. That is, a branch with higher customer traffic will employ more labor, and thus, higher capital. The relationship between labor and capital is assumed to fall in the range of $1800 – 2200 per FTE and this rate is allowed to randomly vary to reflect branch differences.

- Similarly, a positive correlation is assumed between labor and *customer service training* in a given branch, measured in hours. By assuming each unit of labor (FTE) in the division of tellers attracts 50 to 100 hours of training, we account for differences in staff experience and thus, the need for training. The rate for the division of relationship bankers is set higher at 80 to 150 hours per unit of labor in view of the more complicated products/services this division provides regardless of experience. The above rates are allowed to vary randomly across branches and values for customer service training are determined by multiplying rates and labor inputs.

- Since labor is a rough measure of the flow of customers through the division of tellers, we expect the outputs of *number of transactions* and the *number of referrals* to be positively correlated with labor. Each unit of labor is assumed to handle, on average, 30 transactions per day or 6600 transactions per year assuming five workings days in a week and eleven months in a year. Thus, we first randomly generate the *number of transactions per unit of labor* in the range of 6000 - 7200 to accommodate unknown differences across branches. In turn, the *number of referrals* randomly varies between 10 to 40% of the number of transactions, once again, to allow for unknown differences between branch clientele and business mix. This estimate does not include any non-discretionary input of referrals reaching a bank from phone and internet banking services. According to Paradi *et al.* [11, p.99], "…it is through a branch that customers do a large percentage of their more value added banking activities…"

- Finally, the *number of referral sales* in the division of relationship bankers depends on the number of referrals received from the division of tellers. Thus, referral sales are allowed to randomly vary between 50 to 90% of the number of referrals to reflect the differences in successful selling activities of relationship bankers.

2.2 Network slacks-based measure of efficiency (NSBM)

We use DEA-Solver Pro software to execute weighted NSBM assuming variable-returns-to scale (VRS) and non-orientation. In applications of DEA, the slacks-based measure (SBM) by Tone [12] has gradually become the non-radial model of choice. SBM, which belongs to the category of DEA models known as *non-radial*, also has the desirable property of units-invariance. Our choice of variable returns-to-scale accounts for the often different scale of operations found in business organizations. VRS is more desirable because it processes information on returns-to-scale of operations for each DMU (decision-making unit) while avoiding inappropriate extrapolation of performance (see Smith [4], p. 244). VRS is particularly suitable in network DEA which attempts to capture the interactions among multiple divisions within each DMU.

The choice of non-orientation and non-radial modeling further makes frontier efficiency studies more relevant to the world of business. For example, non-orientation ensures the analysis simultaneously captures slacks on both sides as the linear program minimizes inputs and maximizes outputs. Similarly, use of non-radial modeling accommodates the often non-proportional nature of slacks in organizations where production relationships demand different proportions of changes in inputs and outputs. The choice of non-orientation also assists in comparisons following simultaneous data perturbations discussed in section 3.4.

Equations (1) and (2) are adapted from Tone and Tsutsui [10]. Equation (1) defines the objective function for divisional efficiency, and equation (2) receives the optimal input and output slacks from equation (1). Therefore, the overall efficiency of a branch is computed in equation (2). The weighted summation in equation (2) also implies that for the overall NSBM estimate to equal 1, all divisions of a branch must also be efficient. This, in turn, suggests that an overall efficiency estimate of 1 cannot be guaranteed for a given sample because of the multiple divisional production frontiers. In equation (2), in addition to the objective function for a DMU under investigation, we show the respective constraints for divisional weights and intensity values.

$$\rho_j = \frac{1 - \frac{1}{N_j}\left(\sum_{n=1}^{N_j} \frac{s_{no}^{j-*}}{x_{no}^{j}}\right)}{1 + \frac{1}{M_j}\left(\sum_{m=1}^{M_j} \frac{s_{mo}^{j+*}}{y_{mo}^{j}}\right)} \quad (j = 1, \ldots, J) \tag{1}$$

$$\rho_o^* = \min \frac{\sum_{j=1}^{J} w^j \left[1 - \frac{1}{N_j}\left(\sum_{n=1}^{N_j} \frac{s_{no}^{j-}}{x_{no}^{j}}\right)\right]}{\sum_{j=1}^{J} w^j \left[1 + \frac{1}{M_j}\left(\sum_{m=1}^{M_j} \frac{s_{mo}^{j+}}{y_{mo}^{j}}\right)\right]} \tag{2}$$

subject to

$$\sum_{j=1}^{J} w^j = 1, \;\; w^j \geq 0 \; (\forall j), \tag{2a}$$

$$\sum_{k=1}^{K} z_k^j = 1 \; (\forall j), \;\; z_k^j \geq 0 \; (\forall k, j), \text{ and} \tag{2b}$$

$$t^{(j,h)} z^h = t^{(j,h)} z^j . \; (\forall (j,h)), \;\; t_k{}^{(j,h)} = (t_1{}^{(j,h)}, ..., t_K{}^{(j,h)}) \in R^{T^{(j,h)} \times K} \tag{2c}$$

where,

k = a DMU (K = number of DMUs);

j = a division (J = number of divisions);

n = an input (N = number of inputs);

m = an output (M = number of outputs);

w = divisional weight;

s_{no}^{-} = input slack;

s_{mo}^{+} = output slack;

z = intensity;

(j,h) = intermediate product link between division j and division h;

$t^{(j,h)}$ = intermediate product;

$T^{(j,h)}$ = number of intermediate products

Constraint 2(a) indicates non-negative weights for all the divisions that add up to 1, thus allowing separate accounting for the importance of the divisions. Constraint 2(b) indicates non-negative divisional intensities that add up to 1, thus opening the analysis to variable returns-to-scale. Constraint 2(c), introduces free linking where linking activities are discretionary, while maintaining continuity between inputs and outputs and enabling an intermediate product (Tone and Tsutsui [10]).

3. Research design

3.1 Overview

In the proposed network structure of a bank branch (see Figure 2), there are three input variables and one output variable per division, and one intermediate product linking the two divisions. Using the operational branch network performance modeled in this paper adds character and business relevance to sensitivity analysis - qualities that are sometimes lacking in studies with simulated data. Following generation of data for a population of 2000 branches (see next section), we work with a

range of sample sizes N = 20, 40, 80, 160, 320. This range is similar to that reported in Smith [4] where N = 10, 20, 40, 80 and Chapparo *et al.* [3] where N = 10, 20, 40, 80, 160.

We adopt a *nested* sampling approach similar to that discussed in Smith [4]. That is, starting with a randomly selected sample of 320 branches from the population of two-thousand branches, we create the next sample (N = 160) by removing branches 161-320, and the sample after that (N = 80) by removing branches 81-160, and so on. This approach ensures the survival of smaller samples in the larger samples, thus allowing the analyst to monitor a given group, e.g. membership of benchmark group in N=20 followed in larger samples.

3.2 Algorithm for data generation and descriptive population parameters

The following algorithm summarizes the *sequential* steps in generating data for a population of 2000 branches. Data generation incorporates the variable relationships detailed in section 2.1 where all the numbers generated from a given range are random and positive. Numbers in square brackets provide an example of data generation for a mid-sized branch proxied by labor of 10 FTE.

i. Total labor input for the branch:

$$x_{FTE},\ 4...16$$
[10FTE]

ii. Labor for the division of tellers (input):

$$x_{FTE}^{TEL},\ 1.20x_{FTE}^{RB}...1.50x_{FTE}^{RB}\ \ ,where\ \ x_{FTE}^{TEL}+x_{FTE}^{RB}=x_{FTE}$$
[Assuming a factor of 1.3, teller labor equals 5.65FTE]

iii. Labor for the division of relationship bankers (input):

$$x_{FTE}^{RB}=x_{FTE}-x_{FTE}^{TEL}$$
[4.35FTE]

iv. Capital for the division of tellers (input):

$$x_{K}^{TEL}=\ TEL_{FTE}^{K}\ .\ x_{FTE}^{TEL}\ \ ,where\ \ TEL_{FTE}^{K},\ 1800...2200$$
[Assuming a rate of $2000 per FTE, then 2000x5.65=$11,300]

v. Capital for the division of relationship bankers (input):

$$x_K^{RB} = RB_{FTE}^K \ . \ x_{FTE}^{RB} \ , where \ \ RB_{FTE}^K, \ 1800...2200$$

[Assuming a rate of $1900 per FTE, then 1900x4.35=$8,265]

vi. Customer service training for the division of tellers (input):

$$x_{CST}^{TEL} = TEL_{FTE}^{CST} \ . \ x_{FTE}^{TEL} \ , where \ \ TEL_{FTE}^{CST}, \ 50...100$$

[Assuming a rate of 60 hours per FTE, then 60x5.65=339 hours]

vii. Customer service training for the division of relationship bankers (input):

$$x_{CST}^{RB} = RB_{FTE}^{CST} \ . \ x_{FTE}^{RB} \ , where \ \ RB_{FTE}^{CST}, \ 80...150$$

[Assuming a rate of 110 hours per FTE, then 110x4.35=478.5 hours]

viii. Number of transactions for the division of tellers (output):

$$y_{TRAN}^{TEL} = TEL_{FTE}^{TRAN} \ . \ x_{FTE}^{TEL} \ , where \ \ TEL_{FTE}^{TRAN}, \ 6000...7200$$

[Assuming a rate of 6700 transactions per FTE, then 6700x5.65=37,855]

ix. Number of referrals for the division of tellers (output) becomes an input for the division of relationship bankers:

$$y_{REF}^{TEL}, \ 0.10y_{TRAN}^{TEL} ...0.40y_{TRAN}^{TEL} \ , where \ \ y_{REF}^{TEL} = x_{REF}^{RB},$$ i.e. an intermediate product

[Assuming a factor of 0.2, then 0.2x37,855=7571]

x. Number of referral sales for the division of relationship bankers (output):

$$y_{REF\_S}^{RB}, \ 0.50x_{REF}^{RB} ...0.90x_{REF}^{RB}$$

[Assuming a factor of 0.6, then 0.6x7571=4543]

Table 1 shows the descriptive parameters on the population thus generated. In general, variables approach normal distribution and correlations are high. The near-normal distribution can be seen in mean and median values that mostly are close to each other, and low skewness. *Within-input* correlations are higher because of the deterministic algorithm used in generating data and the key role played by the measure of full-time equivalent (FTE).

[Insert Table 1 about here]

3.3 Discrimination across various sample sizes

The middle sample of eighty bank branches with the full complement of variables and divisional weights depicted in Figure 2 is hereafter referred to as the *core model*, and results from the core model in the absence of perturbations are referred to as *baseline results*. Initially, the research design calls for comparisons *across different sample sizes* by monitoring discrimination. This involves observing mean and range of NSBM efficiency estimates, efficient versus inefficient branches, membership of the benchmark group, repositioning of the benchmark branches as sample size grows, and so on.

3.4 Perturbations

In this section, we mainly focus attention on the core model and expose it to a series of data perturbations. As Banker [13, p.155] succinctly states, “The question that is of considerable interest in many applications is whether a set of variables is significant at the margin in characterizing production correspondence between inputs and outputs." Below, we use Banker’s sentiment to start outlining a series of perturbations where we first focus on input/output variables, followed by removal of efficient branches, and switching of divisional weights. Finally, we stack various types of perturbations.

(a) **Deleting relevant inputs**: For example, *capital* or *customer service training* can be deleted across all the divisions.

(b) **Introducing an irrelevant (extraneous) input:** This perturbation involves introducing a variable not found in the modeled production. We randomly generate data for the *average age* of tellers and relationship bankers in the range of 20-40 and 30-55 years old, respectively, assuming more experienced staff would be employed as relationship bankers.

(c) **Simultaneous data perturbations across multiple variables in the core model**: Testing for what Zhu [14] calls the ‘worst-case scenario’, the benchmark branches’ data are adjusted to lower their efficiencies (by simultaneously lowering outputs and raising inputs), and the inefficient branches’ data are adjusted to raise their efficiencies (by simultaneously raising outputs and

lowering inputs). However, the term ‘worst-case scenario’ could be misleading here because we are not predicting what it might entail in practice. Instead, we coin the term ‘convergence scenario’ as the two groups of branches are gently pushed towards each other.

For example, the above perturbation could be justified in the context of a new directive from the bank executive to capture greater market share in a competitive environment while avoiding a blowout of operating costs. Starting from this premise, we specify different increments for each variable based on the following economic rationale. Normally, more inputs would be expended in the division of tellers to create additional referrals. Yet, to accommodate the requirements of simultaneous data perturbations, efficient branches and inefficient branches will be subject to *opposite* changes in their inputs and outputs. We assume that *inefficient* branches will be able to reduce their inputs while increasing their outputs by tapping into their existing inefficiencies. Similarly, as part of the convergence scenario, *efficient* branches will be increasing their inputs while outputs fall because they cannot bring additional effort without using more resources, and they will lose some internal market share to other branches approaching benchmark status. Essentially, we are forcing the inefficient branches towards the benchmark group to see how *responsive* they are, while we push the efficient branches towards the inefficient group in an effort to see how *resilient* they are to simultaneous perturbations.

For the *efficient* branches, in the division of tellers we assume that a 20% rise in labor, a 10% rise in capital, and a 30% rise in customer service training will be accompanied with a 10% drop in the number of transactions and a 5% drop in the number of referrals. In the division of relationship bankers, we assume no changes in labor and capital since a 5% drop in number of referrals generated by the tellers is not going to warrant a substantial change in these inputs. However, we assume the same 30% rise in customer service training for the relationship bankers as they put in an equal effort to tellers even though the number of referral sales also drops by 5%. For the *inefficient* branches, we assume a 20% drop in labor, a 10% drop in capital, and leave customer service training unchanged, while implementing a 10% rise in the number of transactions and a 5% rise in the number of referrals generated in the division of tellers. In the

division of relationship bankers, we assume a drop of 10% in labor and a drop of 5% in capital while leaving customer service training unchanged. Number of referral sales rises by 5%. Simultaneous data perturbations applied to the core model are summarized in Table 2.

[Insert Table 2 about here]

(d) **Deleting benchmark branches (layering)**: We borrow the concept of 'peeling the DEA onion' illustrated in Barr *et al.* [15], and remove the emerging iso-efficiency cohorts until we can identify a *core inefficient cohort*. We stop short of calling iso-efficiency cohorts 'tiers' or 'frontiers' because network SBM does not define a single efficient frontier in the traditional sense; rather, NSBM estimates depend on multiple efficient frontiers formed at the divisional level. Others who have used tiered DEA or layering include Paradi *et al.* [16], Bougnol and Dulá [17], and Gomes *et al.* [18]. Changes in ranking among the members of the core efficient cohort following sequential removal of iso-efficiency cohorts allow us to gauge the sensitivity of inefficient branches to benchmark branches.

(e) **Changing divisional weights:** We start with 0.3 (tellers) and 0.7 (relationship bankers), suggesting that the division of relationship bankers plays a more significant role in revenue generation because of the higher margin products and services it handles. We then swap these weights in recognition of the alternative argument that without the tellers generating referrals, the relationship bankers would mostly be unable to sell.

(f) **Stacking different types of perturbations:** We are also interested in identifying ramifications of executing some of the above perturbations simultaneously. In an illustration using the sample N=80, we combine the simultaneous data perturbations outlined in (c) with deletion of the relevant input variable of *customer service training* and re-setting of divisional weights to 0.7 and 0.3 for tellers and relationship bankers, respectively. Emerging results are then compared against baseline results.

3.5 Generating confidence intervals around NSBM estimates using jackknifing

Horsky and Nelson [19, p.130] remind us that whenever we can make distributional assumptions about error terms, distributions of parameter estimates are known asymptotically and statistical significance tests such as the *t*-test can be designed. Yet, linear programming behind DEA makes no distributional assumptions. According to Simar and Wilson [20], because estimators of the non-parametric frontier are obtained from finite samples, efficiency estimates are sensitive to sampling variations. In fact, it is well established in DEA literature that efficiency estimates are sensitive to sample composition. Sampling variations around the observed efficient frontier may cast doubt on the validity of estimates [21, p.1009].

Employing the sample re-use method of *jackknifing* we show how to test for the potential instability of NSBM estimates using N = 20, 40, and 80. In a similar manner to that of leave-one-out method, a branch is omitted from the original sample and NSBM efficiencies re-estimated for the emerging sample sub-set. The originally omitted branch is then returned before the next branch is removed (thus maintaining the same degrees of freedom throughout re-testing). Network DEA is repeated until all possible samples are exhausted. This is followed by computation of pseudovalues, a jackknife estimator and standard error in order to establish confidence intervals (i.e. upper and lower limits).

The simple yet time consuming jackknifing exercise, that requires manual repeated execution of DEA software, is limited to the three smallest samples given the illustrative nature of the study. This approach makes sense from another perspective as well, namely, given dimensionality, efficiency estimates drawn from smaller samples are more likely to be suspect. If the mean of the estimates from the original sample falls within the confidence intervals generated through re-sampling, then we can reasonably conclude that NSBM estimates are unlikely to be artifacts of the sample used. Paraphrasing Crask and Perreault [22], the essence of jackknifing is to separate the effect of a given branch on the NSBM estimate obtained from the full sample or population. Further details of jackknifing computations are provided in the appendix.

4. Results of sensitivity analyses

4.1 Observations related to sample size

Table 3 provides an overview of model discrimination as sample size rises. Information in the first three rows of the table is in line with established knowledge on traditional DEA. That is, as sample size grows, discrimination improves. Changes in the composition of the benchmark group are also in line with traditional DEA. The low rate of survival of benchmark groups from one sample to the next underscores the relative nature of DEA where new branches outperform branches in the previous sample's benchmark group. In fact, only the efficient branch 181 that is introduced in the second sample survives in the benchmark groups of the following three samples - appearing in 4/5 benchmark groups.

[Insert Table 3 about here]

4.2 Observations related to data perturbations

Table 4 summarizes the results of various perturbations discussed in section 3.4 (except layering which is reported in Table 5). Removing a variable raises degrees of freedom and would normally improve discrimination in traditional DEA. Network SBM results indicate a similar behavior when we omit *capital*, or when we return capital into the model but omit *customer service training* input instead. That is, compared to the baseline results, new NSBM efficiency estimates are over a wider range. However, more noticeable is the absence of any substantial changes in other measures in Table 4. For example, omitting the input of capital has no impact on the benchmark group. Similarly, only one of the benchmark branches leaves the benchmark group when customer service quality is removed. This apparent stability of estimates to deletion of a relevant input is also supported by positive significant rank correlations. On the other hand, introducing an input variable not relevant to branch production, namely, the average age of staff, pushes the results in the opposite direction. That is, the mean estimate rises as the range of estimates contracts, and this moderate loss of discrimination is also evidenced by the tripling of the number of branches retained in the benchmark group.

[Insert Table 4 about here]

Table 4 also depicts the analysis following simultaneous data perturbations where a comparison to baseline results indicates a small loss of discrimination. While there is a noticeable change in the benchmark group, the rank correlation with the baseline efficiency estimates is positive and significant. Membership of benchmarking groups indicates that the previously efficient branch 181 is now inefficient, and branch 536 becomes efficient. Overall, results indicate a model resilient to simultaneous data perturbations across multiple variables in the context of moderate changes imposed.

Regarding the impact of changing divisional weights, the default weights are swapped, where we now use 0.7 for the division of tellers and 0.3 for the division of relationship bankers. Comparing the new results against baseline results reveals a substantial drop in discrimination reflected in a higher mean, a smaller range and a lower coefficient of variation. Nevertheless, composition of the benchmark group remains unchanged, and rank correlation is positive and significant. The latter observation suggests that NSBM can accommodate a range of divisional weights without necessarily compromising the baseline benchmark group.

Returning to Table 4, stacking different types of perturbations (2, 4 and 5) shows the greatest impact on the NSBM efficiency estimates where the analysis substantially loses its ability to differentiate among branches, i.e. a higher mean and a much narrower range of estimates. Membership of the benchmark group that emerges here has no resemblance to the baseline benchmark group. For the first time in the series of perturbations tested here, rank correlation drops to about 0.7, indicating that the estimates move away from the baseline results.

Next we report on findings when benchmark branches are removed from the sample in the core model using layering. Unlike traditional DEA, layering here does not always account for all the branches in a sample because of the nature of NSBM. That is, we reach a point where there are no more efficient branches but a number of inefficient branches still remain. In the sample N=80 the latter group (layer 8) consists of fifty-six inefficient branches, and we dub this group the *core inefficient cohort.* In Table 5, as layering creates a smaller sample at each step, there is evidence of some loss of discrimination in the steadily rising mean and median of estimates corresponding to the *core inefficient cohort*. Yet, significant rank correlations capturing changes between consecutive

layering imply resilience of inefficient branches against removal of benchmark branches from the sample. That is, there is a core group of inefficient branches whose measure of relative performance is not substantially affected by omission of benchmark branches in the sample. Management may feel more confident focusing improvement activities on this core inefficient cohort in the knowledge that their relative ranking is not overly tied to one particular group of efficient branches.

[Insert Table 5 about here]

We re-compute rank correlations in Table 5 using a more strict approach by following changes in estimates that are separated by three layers, i.e. the core inefficient cohort in layer 1 versus in layer 4, in layer 2 versus in layer 5, and so on. Rank correlations reported in the last row in Table 5 are noticeably lower compared to the previous row of correlations based on consecutive layers. Nevertheless, these correlations are still statistically significant at the 1% level of confidence, thus adding further confidence to our earlier conclusion about the resilience of branches in the core inefficient cohort.

Before leaving layering, we also briefly report on our additional layering attempts using other samples in this study. Samples N=20 and N=40 behave similar to layering encountered in traditional DEA where repeated removal of benchmark branches exhausts the samples after six and eight layers, respectively. This is contradicted by our findings from the larger samples of N=80, N=160 and N=320 which show seven, eight and sixteen benchmark cohorts, respectively, but fail to exhaust the samples. In these larger samples, 70-85% of the sample remains inefficient after layering, thus opening the way for management to identify a core inefficient cohort in the manner we demonstrated with N=80. Further details are available upon request.

4.3 Observations related to jackknifing

Finally, we offer some jackknifing results on the sensitivity of network efficiency estimates to sampling variations. In section 3.5 we presented some arguments in favor of generating confidence intervals around estimates. For illustrative purposes, this exercise is limited to the three smallest samples in the study where estimates are more likely to be affected by sampling variations. For the core model (i.e., N=80), the procedure detailed in the appendix results in upper and lower limits of

0.4342and 0.5947, respectively. Since the mean NSBM estimate (0.5824) from the core model falls within this interval, we conclude that there is no reason to be alarmed about the observed estimates overall being artifacts of sampling. For N=40, the confidence interval is 0.3703 – 0.5950, and for N=20 the interval is 0.4202 – 0.7594, where the respective sample mean NSBM estimates are 0.6046 and 0.7227, i.e. N=40 sample mean falls just outside the interval. For N=40, only when we recalculate the confidence interval at the higher confidence level of 98% do we manage to bring the sample mean into the interval (new interval becomes 0.3479 – 0.6174). In summary, two out of the three samples' mean estimates appear not to suffer from sampling variations at the 95% confidence level.

5. Discussion

Given the limited space available in journal articles, it is not feasible to attend to all the possible perturbations. Furthermore, the current study is designed to open the way for sensitivity analysis in network DEA, rather than parade as a definitive paper. Hopefully, the paper will provoke further innovative thinking in this exciting field. Therefore, we briefly pause to acknowledge some potential extensions of the current study on sensitivity of network DEA. For example, the perturbations used in this study can be replicated for alternative network structures (see Table 2 and the brief discussion in Kao and Hwang [23, p.442]), as well as hybrid models that combine network DEA with other approaches, e.g. see Yu and Lin's [24] multi-activity network DEA. Other future directions include testing the sensitivity of network DEA through more sophisticated methods designed to provide statistical inferences on efficiency estimates and coding such methods into commercially available software, thus increasing accessibility and utilization. For example, while the bootstrapping method by Simar and Wilson has been coded into FEAR which runs in *R* code, the average DEA researcher is not sufficiently skilled in coding to extend the basic examples provided therein to such methods as NSBM. Such bottlenecks continue to impede take up of bootstrapping.

Since 1980s, one of the criticisms leveled at DEA is that the technique is deterministic and does not offer the range of diagnostics that often accompany econometric methods of frontier estimation (see various studies such as Banker [13], and Simar and Wilson [20] and [25], that respond

to this criticism). That is, while traditional DEA reports DMU (decision-making unit) performance in unique levels of inefficiency, it fails to account for the uncertainty that may be embedded in such estimates. Where it is not possible to separate random noise, DEA incorrectly accounts for the latter as part of inefficiency. Management scientists concerned with this can employ such methods as bootstrapping that can help generate confidence intervals for every estimate and also assist in testing hypotheses on group comparisons. Nevertheless, despite examples of bootstrapping in DEA methodology literature, majority of papers that apply DEA still do not use bootstrapping for inferring uncertainty in efficiency estimates – a point also conceded by Odeck [21].[1] In the current study, we have used jackknifing as an easy-to-implement approach to generate confidence intervals around sample mean NSBM estimates.

A further possible extension of sensitivity analysis of network DEA may include estimation of true efficiencies once mathematical challenges introduced by a network structure are overcome. Unfortunately, most of the existing sensitivity studies that attempt to generate true efficiencies use a Cobb-Douglas type function (see Cobb and Douglas [27]) and make many strict distributional assumptions in generating variables that comprise a linear production function. Furthermore, distributional assumptions are often glossed over by reference to other methodology papers, rather than justified based on observations about data from the world of business. Thus, the relevance of findings from such studies to applications of DEA remains untested. This is one of the reasons the current paper maps data generation in a more deterministic manner by targeting the production processes expected in a real-life bank branch.

Another limitation of sensitivity studies that benchmark DEA estimates against true efficiency estimates is the aggregation of outputs into a single variable to accommodate Cobb-Douglas type functions (an exception is the approach in Ruggiero [5]). Relying on a pre-defined model of an otherwise unobserved production function makes many users of DEA somewhat uneasy because they

[1] For a notable exception, see an application of Simar and Wilson's fine work on bootstrapping in Tortosa-Ausina *et al.* [26].

are drawn to DEA in the first place because of the scarcity of assumptions that enable DEA. The fact that benchmarking against true efficiencies is not reported in applications of DEA is testimony to this observation, as well as the difficulty of formulating convincing production functions that can be used in conjunction with Monte Carlo simulation and bootstrapping. In the current study, which constitutes the first systematic investigation of sensitivity of network DEA estimates, generating true efficiencies would be even more complicated given multiple inputs, intermediate products, multiple outputs, and an efficient NSBM frontier that cannot be defined in a similar fashion to traditional DEA. Furthermore, bootstrapping cannot be reliably attempted unless the data generating process for non-radial, non-oriented, variable returns-to-scale, weighted NSBM is well defined. In short, many challenges await future research into this field.

We ought to also mention that recent years have witnessed the addition of *robust optimization* to the mathematical techniques designed to investigate the sensitivity of DEA estimates. Robust optimization involves modeling uncertainty into data by constraining input output parameters with additional constraints based on the worst-case solution. Readers interested in pursuing this line of inquiry are referred to Bertsimas *et al.* [28] and Shokouhi *et al.* [29] among others.

Finally, there are some obvious limitations of executing one perturbation at a time. We have illustrated a case where multiple perturbations are administered simultaneously – once again – in an effort to create an environment of experimentation that is likely to better reflect the real world of business. Future studies on robustness of network DEA may re-visit the possibility of stacking perturbations in other combinations.

6. Conclusions

This paper provides a systematic investigation of the sensitivity of network data envelopment analysis following various types of data perturbations and represents the first such attempt in literature. The primary objectives are, (a) to generate insight to robustness of network DEA, (b) to achieve objective (a) with data generated in an intuitive production environment that mimics real-life business, (c) to raise the confidence in NDEA estimates with the intention of encouraging a wider use

in organizational decision-making, and (d) provide an open discussion of the challenges that lie ahead with a view to attracting the attention of others to the field.

Overall, all rank correlations are positive and significant when post-perturbation results are compared against baseline results, which can be construed as preliminary empirical evidence on stability. Key findings following various data perturbations to the popular DEA model of network slacks-based measure (NSBM) can be summarized as, (a) increasing sample size improves discrimination, (b) removing a relevant input improves discrimination but has either no or little effect on the composition of the benchmark group, (c) introducing an extraneous input leads to a moderate loss of discrimination that creates a much enlarged benchmark group, (d) simultaneously adjusting inputs and outputs in opposite directions for inefficient versus efficient branches indicates a mainly stable NSBM, (e) swapping divisional weights produces a substantial drop in discrimination while the benchmark group remains identical, (f) stacking perturbations has the greatest impact on efficiency estimates with substantial loss of discrimination, emergence of a different benchmark group and a much lower rank correlation, and (g) layering results suggest that the core inefficient cohort is resilient against omission of benchmark branches.

There are a number of managerial implications that can be construed from the empirical findings. For example, starting with the selection of mathematical modeling options, the choice of non-orientation and non-radial approaches to efficiency measurement makes the study design more relevant to business applications. That is, non-orientation ensures the analysis simultaneously captures input as well as output slacks, and non-radial modeling accounts for the non-proportional nature of slacks often found in organizations where production relationships demand that inputs contract and outputs expand in different proportions. Similarly, selecting a larger sample size is expected to bring increased discriminatory power to the analysis, thus helping management better identify where the potential improvements lie.

Focusing on various perturbations illustrated, for example, introducing an input variable not relevant to branch production leads to a loss of discrimination in analysis and is a reminder to

management of the need to select variables with care. The example of simultaneous data perturbations where inefficient and efficient groups of branches are pushed towards each other allows management to identify the effort required in capturing greater market share while keeping an eye on resources expended. Similarly, experimenting with different divisional weights opens the door for further managerial insight to varying the role of divisions in generating value for the bank branch unit. This flexibility in scenario analysis may also help executive management better sell the idea of branch performance modeling to line management by showing how each party's viewpoint impacts overall branch performance.

Stacking different types of perturbations is also conducive to what-if analyses where management can test various scenarios they anticipate about the future of branches' operating environment. Essentially, such stacked perturbations can be used to see what is likely to happen to branch performance given anticipated conditions that may better reflect the more complex operating environment businesses deal with in practice. Finally, identifying the *core inefficient cohort* of branches whose measure of relative performance is not overly tied to a particular group of efficient branches may provide management a focus on improvement activities for those branches that are most likely to benefit.

References

[1] Färe R, Grosskopf S. Intertemporal Production Frontiers: With Dynamic DEA. Boston: Kluwer Academic Publishers; 1996.

[2] Färe R, Grosskopf S. Network DEA. *Socio. Econ. Plan. Sci.* 2000; 34: 35-49.

[3] Chapparo FP, Salinas-Jimenez J, Smith P. On the quality of the data envelopment analysis. *J. Oper. Res. Soc.* 1999; 50: 636-644.

[4] Smith P. Model misspecification in Data Envelopment Analysis. *Ann. Oper. Res.* 1997; 73: 233-252.

[5] Ruggiero J. A new approach for technical efficiency estimation in multiple output production. *Eur. J. Oper. Res.* 1998; 111: 369-380.

[6] Seiford L, Zhu J. Sensitivity and stability of the classifications of returns to scale in data envelopment analysis. *J. Productivity Anal.* 1999; 12: 55-75.

[7] Galagedera D, Silvapulle P. Experimental evidence on robustness of data envelopment analysis. *J. Oper. Res. Soc.* 2003; 54: 654-660.

[8] Avkiran NK. Stability and integrity tests in data envelopment analysis. *Socio. Econ. Plan. Sci.* 2007; 41: 224-234.

[9] Avkiran NK. Opening the black-box of efficiency analysis: An illustration with UAE banks. *OMEGA – Int. J. Manage. S.* 2009; 37: 930-941.

[10] Tone K, Tsutsui M. Network DEA: A slacks-based measure approach. *Eur. J. Oper. Res.* 2009; 197: 243-252.

[11] Paradi J, Rouatt S, Zhu H. Two-stage evaluation of bank branch efficiency using data envelopment analysis. *OMEGA – Int. J. Manage. S.* 2011; 39: 99-109.

[12] Tone K. A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* 2001; 130: 498-509.

[13] Banker RD. Hypothesis tests using data envelopment analysis. *J. Productivity Anal.* 1996; 7: 139-159.

[14] Zhu J. Super-efficiency and DEA sensitivity analysis. *Eur. J. Oper. Res.* 2001; 129: 443–55.

[15] Barr RS, Durchholz ML, Seiford L. Peeling the DEA onion: Layering and rank-ordering DMUs using tiered DEA. Southern Methodist University Technical Report, 1994, revised in 2000, http://faculty.smu.edu/barr/pubs/.

[16] Paradi J, Asmild M, Simak PC. Using DEA and worst practice DEA in credit risk evaluation. *J. Productivity Anal.* 2004; 21: 153-165.

[17] Bougnol ML, Dulá JH. Validating DEA as a ranking tool: An application of DEA to assess performance in higher education. *Ann. Oper. Res.* 2006; 145: 339-365.

[18] Gomes EG, de Mello JCCBS, e Souza GS, Meza LA, Mangabeira JAC. Efficiency and sustainability assessment for a group of farmers in the Brazilian Amazon. *Ann. Oper. Res.* 2009; 169: 181.

[19] Horsky D, Nelson P. Testing the statistical significance of linear programming estimators. *Management Sci.* 2006; 52: 128-135.

[20] Simar L, Wilson PW. Sensitivity analysis of efficiency sores: How to bootstrap in nonparametric frontier models. *Management Sci.* 1998; 44: 49-61.

[21] Odeck J. Statistical precision of DEA and Malmquist indices: A bootstrap application to Norwegian grain producers. *OMEGA – Int. J. Manage. S.* 2009; 37: 1007-1017.

[22] Crask MR, Perreault WD. Validation of discriminant analysis in marketing research. *J. Marketing Res.* 1977; 14: 60-68.

[23] Kao C, Hwang S-N. Efficiency measurement for network systems: IT impact on firm performance. *Decis. Support. Syst* . 2010; 48: 437-446.

[24] Yu M-M, Lin ETJ. Efficiency and effectiveness in railway performance using a multi-activity network DEA model. *OMEGA – Int. J. Manage. S.* 2008; 36: 1005-1017.

[25] Simar L, Wilson PW. Statistical inference in nonparametric frontier models: The state of the art. *J. Productivity Anal.* 2000; 13: 49-78.

[26] Tortosa-Ausina E, Grifell-Tatjé E, Armero C, Conesa D. Sensitivity analysis of efficiency and Malmquist productivity indices: An application to Spanish savings banks. *Eur. J. Oper. Res.* 2008; 184: 1062-1084.

[27] Cobb CW, Douglas PH. A theory of production. *Am. Econ. Rev.* 1928; 18: 139-165.

[28] Bertsimas D, Pachamanova D, Sim M. Robust linear optimization under general norms. *Oper. Res. Lett.* 2004; 32: 510-516.

[29] Shokouhi AH, Hatami-Marbini A, Tavana M, Saati S. A robust optimization approach for imprecise data envelopment analysis. *Comput. Ind. Eng.* 2010; 59: 387-397.

Figure 1: An example of a black box production under traditional DEA

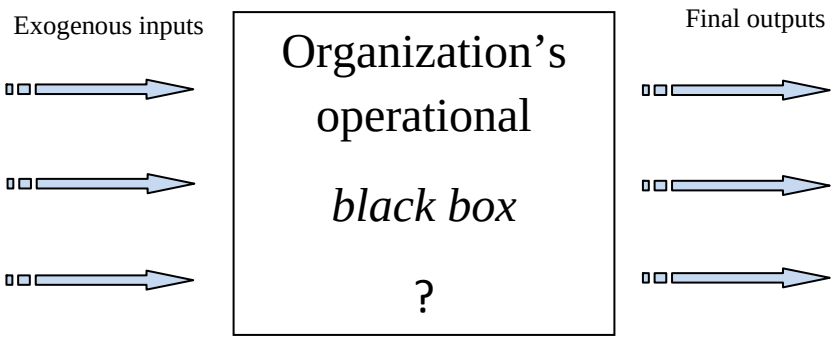
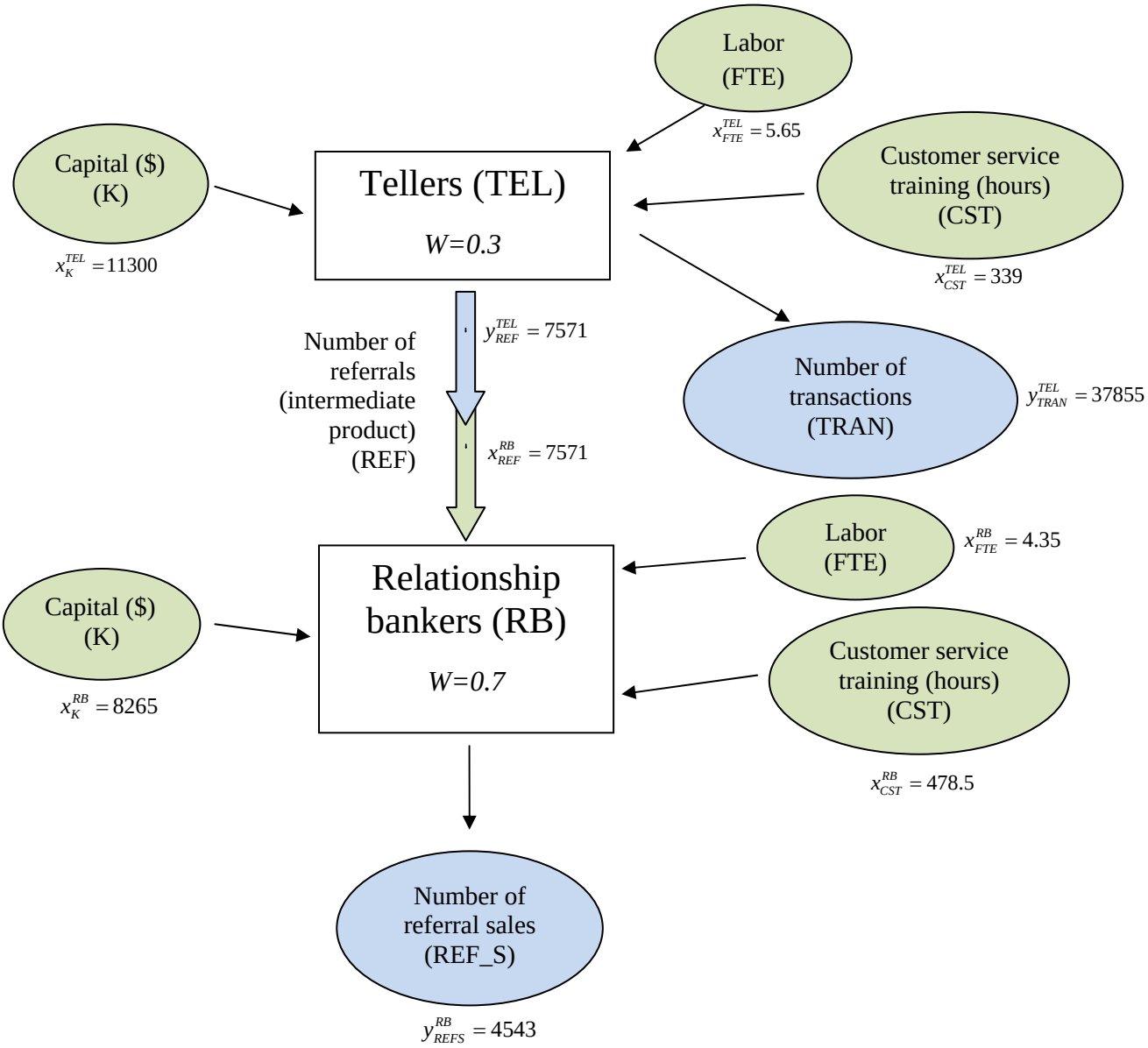


Figure 2: A weighted, two-division network structure of a basic bank branch production



Notes: The above diagram focuses on the main service providers. Values corresponding to the data generation example in section 3.2 are shown next to variables. The divisional weights depicted in Figure 2 assume that relationship bankers play a greater role in revenue generation.

FTE: Full-time equivalent – a unit that overcomes the difficulty of measuring labor when different mixes of full-time and part-time appointments co-exist.

Table 1: Descriptive parameters of the population of 2000 branches and correlations

Panel A: Descriptive parameters

	Inputs and outputs in the division of tellers				Inputs and outputs in the division of relationship bankers				Intermediate product
	TEL-FTE	*TEL-K*	*TEL-CST*	TEL-TRAN	*RB-FTE*	*RB-K*	*RB-CST*	RB-REF_S	**TEL-REF**
Mean	5.74	11,472.74	429.90	37,930.49	4.25	8,499.48	487.50	6,663.85	9,549.39
Median	5.70	11,423.42	416.25	37,567.91	4.20	8,458.35	471.54	5,957.78	8,667.26
SD	2.02	4,076.76	175.40	13,492.11	1.49	3,030.27	195.12	3,627.10	4,886.27
CV	0.35	0.36	0.41	0.36	0.35	0.36	0.40	0.54	0.51
Maximum	9.53	20,262.44	932.40	66,921.93	7.25	15,447.69	1,037.00	20,534.51	25,137.08
Minimum	2.20	4,034.77	113.52	13,848.03	1.61	3,026.04	128.93	1,021.23	1,579.15
Skewness	0.03	0.06	0.44	0.08	0.05	0.09	0.42	0.93	0.70
Kurtosis	-1.22	-1.13	-0.52	-1.13	-1.18	-1.11	-0.53	0.57	-0.16

Panel B: Bivariate correlations

	TEL-FTE	*TEL-K*	*TEL-CST*	*RB-FTE*	*RB-K*	*RB-CST*	TEL-TRAN	RB-REF_S	**TEL-REF**
TEL-FTE	1.00								
TEL-K	*0.99*	1.00							
TEL-CST	*0.86*	*0.84*	1.00						
RB-FTE	*0.98*	*0.97*	*0.84*	1.00					
RB-K	*0.97*	*0.95*	*0.83*	*0.99*	1.00				
RB-CST	*0.86*	*0.84*	*0.73*	*0.88*	*0.87*	1.00			
TEL-TRAN	0.99	0.97	0.85	0.97	0.96	0.85	1.00		
RB-REF_S	0.64	0.63	0.54	0.62	0.62	0.57	0.65	1.00	
TEL-REF	0.69	0.68	0.58	0.67	0.67	0.61	0.70	0.94	1.00

Notes: SD, standard deviation; CV, coefficient of variation (SD/mean); Input variables are depicted in *italics* and the intermediate product is in bold; TEL, division of tellers; FTE, full-time equivalent; K, capital; CST, customer service training; RB, division of relationship bankers; TRAN, number of transactions; REF_S, number of referral sales; REF, number of referrals.

Table 2: Simultaneous data perturbations applied to the core model

	Percent perturbation in *efficient* branch divisions		Percent perturbation in *inefficient* branch divisions	
	Tellers ⇒	Relationship bankers	Tellers ⇒	Relationship bankers
Inputs				
Labor	20% rise	0%	20% drop	10% drop
Capital	10% rise	0%	10% drop	5% drop
Customer service training	30% rise	30% rise	0%	0%
Outputs				
Number of transactions	10% drop	n/a	10% rise	n/a
Number of referrals	5% drop	n/a	5% rise	n/a
Number of referral sales	n/a	5% drop	n/a	5% rise

Table 3: Discrimination across nested samples [a]

Sample size	*N = 20*	*N = 40*	*N = 80*	*N = 160*	*N = 320*
Mean NSBM estimate	0.7227	0.6046	0.5824	0.5702	0.5509
Range of NSBM estimates	0.2847 - 1	0.2743 - 1	0.2411 - 1	0.2156 - 1	0.1859 - 1
Proportion of NSBM *efficient* branches [b]	6/20 = 30.00%	5/40 = 12.50%	3/80 = 3.75%	4/160 = 2.50%	7/320 = 2.19%
Changing composition of the benchmark group [c]	n/a	2/5 = 40%	1/3 = 33%	1/4 = 25%	4/7 = 57%

[a] 'Nested' means N=40 contains N=20, N=80 contains N=40, and so on

[b] Ratio of the number of NSBM *efficient* branches to the sample size

[c] Ratio of previous sample's benchmark group represented on the current sample's benchmark group

Table 4: Impact of various perturbations (N=80)

	Baseline results [a]	(1) Without capital	(2) Without customer service training	(3) Introducing average age of staff	(4) Simultaneous perturbations	(5) Swapping divisional weights	(2) (4) (5) Stacked perturbations
Mean NSBM estimate	0.5824	0.5673	0.5578	0.6509	0.6214	0.7266	0.7189
Median NSBM estimate	0.5595	0.5370	0.5449	0.6489	0.6119	0.7361	0.7322
Range of NSBM estimates	0.2411 – 1	0.2343 – 1	0.2154 – 1	0.2648 – 1	0.2584 – 1	0.4290 – 1	0.4420 – 1
Coefficient of variation [b]	0.3455	0.3603	0.3599	0.3329	0.3334	0.1953	0.2018
Skewness of NSBM estimates	0.4120	0.4483	0.3958	0.1325	0.1893	-0.1033	-0.1365
Proportion of NSBM *efficient* branches [c]	3/80 = 3.75%	3/80 = 3.75%	2/80 = 2.50%	9/80 = 11.25%	4/80 = 5%	3/80 = 3.75%	1/80 = 1.25%
Membership of the benchmark group (branch ID numbers)	181, 436, 490	181, 436, 490	181, 436	40, 116, 181, 236, 283, 294, 322, 436, 490	294, 436, 490, 536	181, 436, 490	294
Kendall's *tau-b* rank correlation with baseline results (two-tailed) [d]	n/a	0.921	0.848	0.755	0.837	0.828	0.703

[a] The sample N=80 with the full complement of variables and divisional weights shown in Figure 2 is the core model, and results obtained from the core model in the absence of perturbations are known as baseline results. [b] Coefficient of variation is the ratio of standard deviation to mean. [c] Ratio of the number of NSBM *efficient* branches to the sample size. [d] All rank correlations are statistically significant at the 1% level. In literature, tests often report Spearman or Kendall's tau-b rank correlations. Initially, we obtained rank correlations based on both tests but only report Kendall's. All of the Spearman correlations were higher than Kendall's. Thus, those with a more conservative nature may prefer Kendall's tau-b over Spearman's rank correlations.

Table 5: Impact on the core inefficient cohort of deleting benchmark branches through layering (N=80)

	Layer 1 [a]	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7	Layer 8
Mean NSBM estimate	0.4859	0.5289	0.5509	0.5528	0.5734	0.5888	0.6527	0.6602
Median NSBM estimate	0.4895	0.5337	0.5468	0.5518	0.5808	0.5888	0.6389	0.6580
Range of NSBM estimates	0.2411-0.7811	0.2652-0.8500	0.2652-0.9562	0.2652-0.9624	0.3048-0.9624	0.3119-0.9698	0.3122-0.9698	0.2987-0.9819
Coefficient of variation [b]	0.2742	0.2686	0.2894	0.2909	0.2840	0.2895	0.2827	0.2899
Skewness of NSBM estimates	0.0958	0.0677	0.3425	0.3495	0.3097	0.4142	0.1819	0.0592
Kendall's *tau-b* two-tailed rank correlations [c]								
between consecutive layers	n/a	0.869	0.916	0.990	0.975	0.930	0.800	0.939
three layers apart	n/a	n/a	n/a	0.813	0.901	0.912	0.751	0.737

[a] Progress of the layers indicates repositioning of branches in the core inefficient cohort following removal of benchmark branches, i.e. iso-efficiency cohorts.

[b] Coefficient of variation is the ratio of standard deviation to mean

[c] All rank correlations are statistically significant at the 1% level.

Appendix: Jackknifing

The jackknifing procedure begins with the generation of network slacks-based measure (NSBM) efficiency estimates from a sample where one of the bank branches has been omitted. Next, the omitted branch is returned and another branch taken out before NSBM is repeated on the new sample or sub-set, and so on. For each sub-set, the mean NSBM estimate is then substituted into equation (1) to obtain the so-called *pseudovalue*, J_i , (Crask and Perreault [22]):

$$J_i = k\,\bar{x}^{\rho} - (k-1)\,\bar{x}_i^{\rho} \quad , \qquad i = 1,...,k \tag{1}$$

where,

k = number of sub-sets (equals sample size N)

$\bar{x}^{\rho}$ = mean NSBM estimate for the original sample prior to re-sampling

$\bar{x}_i^{\rho}$ = mean NSBM estimate for sub-set i following re-sampling

Assuming *k* different pseudovalues are independent, identically distributed random variables (Crask and Perreault [22]), then the mean of pseudovalues becomes the *jackknife estimator*, $\hat{J}$:

$$\hat{J} = \frac{\left(\sum_{i=1}^{k} J_i\right)}{k} \tag{2}$$

Equation (3) provides the *standard error* of the jackknife estimator:

$$S.E._{\hat{J}} = \frac{s}{\sqrt{n}} \tag{3}$$

where,

n = sub-set size (i.e. N-1)

s = standard deviation of pseudovalues

Equation (4) delivers the *confidence interval* around the jackknife estimator:

$$\hat{J} \pm \; t\,critical \times (S.E) \tag{4}$$

where the *critical t* value for two-tailed, 95 per cent confidence level and N-1 degrees of freedom is used. If the mean NSBM estimate for the sample, $\bar{x}^{\rho}$, falls within the confidence interval thus determined, then we would be less concerned with sampling variations.