Marketing Science Institute Special Report 10-200

DEA with Econometrically Estimated Individual Coefficients: A Pharmaceutical Sales Force Application

André Bielecki and Sönke Albers

DEA with econometrically estimated individual coefficients: A pharmaceutical sales force application

André Bielecki

Sönke Albers *

February 2010

Department for Innovation, New Media, and Marketing
Christian-Albrechts-University at Kiel
24098 Kiel
Germany
+49 431 880 -4795 (André Bielecki)
-1541 (Sönke Albers)
+49 431 880 -1166 (fax)
bielecki@bwl.uni-kiel.de (email)
albers@bwl.uni-kiel.de (email)

*Corresponding author.

Abstract

In business practice, the efficiency benchmarking technique DEA has been met with high approval. However, its mathematical programming and cross-sectional data framework have several drawbacks. DEA weights can sometimes be technically valid but practically ill-specified and implausible. Parameter significance or model fit indicators such as R^2 cannot be obtained. Another drawback is that an increasing number of variables leads to an increasing number of efficient units. This may lead to wrong conclusions. We propose a model based on multiple observations per unit that combines econometric estimation of individual coefficients with DEA evaluation techniques. Individually estimated coefficients are used as factor weights for efficiency evaluation operations comparable to the DEA method. The model maintains core DEA features and provides valid individual weights, a reduced number of efficient units, parameter significance, and statistical fit as additional advantages. An application for the sales force of a pharmaceutical company illustrates how this method changes the benchmarking results and increases the potential for efficiency improvement.

Keywords: Data Envelopment Analysis, Efficiency Measurement, Sales Management

1. Introduction

In management practice, benchmarking is an important technique for performance improvement. The goal is that lower performing units learn from higher performing units to improve overall performance. This requires an identification of both inefficient and efficient units. In practice, we find two different procedures: First, units are assessed in a subjective and qualitative way (e.g., based on discussions and interviews). Second, a more objective and quantitative analysis based on data is used. In the area of data driven benchmarking, Data Envelopment Analysis (DEA) is the most frequently used benchmarking technique and has been widely accepted. DEA was first published more than thirty years ago by Charnes, Cooper, and Rhodes (1978) and has become well-known. We focus on this data driven understanding of benchmarking.

DEA has been applied to a wide range of relative efficiency benchmarking contexts, such as the evaluation of salespersons (Boles, Donthu, and Lohtia 1995; Mahajan 1991), retail stores (Rhonda et al. 1998), or the benchmarking of marketing productivity (Donthu, Hershberger, and Osmonbekov 2005), but it has a distinct focus on banking branches (e.g., Sherman and Gold 1985, Ferrier and Lovell 1990; Oral and Yolalan 1990; Giokas and Vassiloglou 1991; Aly et al. 1990).

Technically, DEA determines the efficiency of each decision making unit (DMU) (e.g., each salesperson or banking branch) relative to all other units so that the model provides a relative efficiency ranking of all observed DMUs. The top performing units have a maximum efficiency score of 100% and are used as benchmarks for less efficient DMUs, which have an efficiency score of less than 100%. Comparing inefficient DMUs with their individual benchmarks, challenging performance objectives can be formulated, such as sales objectives for inefficient salespersons. If management is able to supervise the inefficient DMUs to improve their work such that they meet the performance levels of their peers, the overall output (e.g., sales volume) can be increased.

To derive an individual unit's efficiency, DEA determines individualized production functions for each observed unit. The efficiency is based on the individual transformation of input to output factors such as how a sales force transforms customer visits and marketing events into sales volume. Benchmark units for each inefficient DMU face similar working conditions (e.g., similar input factor characteristics compared to the observed unit) so that they can give target-specific efficiency improvements. DEA uses linear programming and available cross-sectional data to determine DMU-specific weights for each of the production function's input factors so that the ratio of output to input is maximized. In the above-mentioned sales force example, DEA determines the weights for the combination of the two input factors (customer visits and marketing events) so that the ratio of the sales volume (output) to the two weighted input factors is maximized. By simultaneously computing the factor weights and maximizing the ratio of output (sales volume) to input (customer visits and marketing events), each salespersons' weights are determined as much in favor of the respective salesperson as possible. This ensures that benchmarking partners are identified who, for the same production function, achieve better results. If another DMU achieves an efficiency score higher than that of the observed unit, using the observed unit's weights, the better performing DMU is used as a benchmark for the observed unit. If the observed (inefficient) DMU improves its efficiency by achieving the output level of its efficient benchmark (e.g., a higher sales volume), it achieves full efficiency and raises the overall output level (e.g., sales volume).

Because DEA uses linear programming and cross-sectional data for the calculation of factor weights, the model lacks some advantages of econometric estimation offered with panel data. Based on single observations per unit, conventional DEA provides idiosyncratic parameter values of a production function that maximize the efficiency of the observed unit relative to other units with the help of linear programming. They do not maximize the overall model fit as econometrically derived coefficients of production or response functions do. This can lead to ill-specified parameter values that do not generate good predictions of output for varied input factors and can result in implausible output objectives that imply a low predictive validity of conventional DEA. This ill-specification problem of factor weights is increased by a lack of significance information for the weights and goodness-of-fit information, like an R^2, which both cannot be provided by linear programming. The ill-specified parameter values can also lead to units being falsely identified as benchmarks, which can invalidate the complete benchmarking process when false benchmarks give biased orientation for other DMUs.

The ill-specification problem comes with an additional drawback of the conventional DEA. The analysis frequently results in a large number of efficient units, occasionally approaching 50% of the observed DMUs. In this respect, a large number of input and output factors makes this drawback even worse because it reduces DEA's ability to find similar units for comparison. Non-comparable units reduce the ability to discriminate between efficient and inefficient DMUs (Dyson et al. 2001).

Today, econometric estimation approaches based on multiple observations per unit are available that provide individual parameter estimation with the help of computer simulations, such as hierarchical Bayes (Rossi and Allenby 2003), as an alternative to linear programming. These approaches were not fully developed thirty years ago and can reduce the aforementioned DEA drawbacks. Therefore, we propose a combination of econometric estimation of individual parameters and DEA techniques for efficiency evaluation that can preserve most advantages of the conventional DEA and also provide additional features from the econometric estimation. Our approach provides more accurate parameter values (as indicated by a holdout validation) and no implausible output objectives. Both lead to an improved predictive validity and more potential for efficiency improvement through sharper discrimination between efficient and less efficient DMUs.

Econometrically estimating idiosyncratic coefficients of a production or response function requires panel data whereas the conventional DEA only needs cross-sectional data, because individual coefficients are determined by comparing one unit with other units. In marketing and sales management, data are usually characterized by multiple observations per unit. Therefore, panel data requirements are not a serious problem.

Technically, our recommended DEA with econometrically derived weights suggests the following improvements. We propose to use hierarchical Bayes (HB) to determine econometrically estimated individual coefficients for DEA production functions. Of course, other random parameter estimation techniques (e.g., maximum simulated likelihood) may also be used. The coefficients do not maximize the relative efficiency of the observed DMU like DEA parameters do, but rather maximize the overall model fit, which enriches the information provided by the weights and allows good predictions of output for varied input factors. In addition to the superior estimator, the significance of the estimated weights and the goodness-of-fit, as provided by an R^2, can be assessed.

Along with the econometric estimation, the richer data based on multiple observations per unit avoid the effect of the conventional DEA that an increased number of variables generates an increased number of efficient units (Jenkins and Anderson 2001) and thus improves the discriminatory power of the model. While in conventional DEA, the more variables used for output generation, the higher the probability that the observed unit has a unique (sub-) combination of factors without suitable benchmarks, our recommended data structure provides richer information, even for rarely used variable ranges.

The econometric estimation reduces the ill-specification of factor weights and, in combination with the multiple observations per unit, provides more discriminatory power. The output objectives derived from the econometric estimation are more precise, and the higher discriminatory power leads to less efficient units, which, compared to the conventional DEA, results in higher potential for efficiency improvement.

The remainder of the paper is organized as follows. Section 2 gives a more detailed review on DEA restrictions related to our model. Section 3 explains the improved relative efficiency modeling. Section 4 illustrates the model using a pharmaceutical sales force application and compares the results to a conventional DEA. Section 5 points out the main managerial implications. Section 6 presents the conclusions and implications for further research.

2. Literature Review

A well-written introduction to DEA can be found in Cooper, Seiford, and Tone (2006), Zhu (2003), or as a short introduction in Avkiran (1999). In summary, DEA offers the following four benchmarking features for practical applications (Charnes, Cooper, and Rhodes 1978):

1. DEA determines a ranking of DMUs (e.g., salespersons) according to their relative efficiency based on input to output transformations.
2. DEA computes weights for various input factors indicating the contribution of the observed factor to the output generation process.

3. DEA identifies benchmark units for all DMUs that are not fully efficient. Benchmark units are assumed to have a similar production structure as the observed DMU, thus allowing comparison between the two.
4. DEA calculates a value for the influence of certain benchmarking units on the observed DMU, which makes it easier to determine which benchmark has the highest impact on the observed DMU. The differing influences of benchmark units on the observed DMU provide additional information for a prioritization of tasks for future efficiency improvements.

Despite its wide range of applications, DEA has some important drawbacks.

The linear programming approach is based on single observations per unit (cross-sectional data). If multiple observations per unit are available (panel data), DEA uses them to analyze the development over time with, for example, windows analysis (Charnes et al. 1985a; Charnes et al. 1985b; Charnes et al. 1994), but multiple observations are not used to improve the estimation procedure as is. Hence, conventional DEA only uses a single observation per DMU for determining individual coefficients for the production (response) function of each DMU by comparing the efficiency of a certain DMU with all the others, respectively. The individually estimated production (response) coefficients are estimated as favorably as possible for the observed DMU while the model maximizes the respective unit's relative efficiency, which is in contrast to a maximization of the overall model fit as in econometric models. This procedure can occasionally result in technically valid but practically ill-specified, and thus implausible, production functions that do not allow good predictions of output for varied input factors, resulting in low predictive validity. Using linear programming to determine individual coefficients, conventional DEA weights cannot be evaluated on grounds of significance information. The model also lacks goodness-of-fit information, like R^2 (Cooper, Seiford, and Tone 2006). An additional problem appears while maximizing individual relative efficiency. In the case of extreme factor utilizations, or rare combinations of input factors, maximization of individual relative efficiency in conventional DEA may result in a DMU being rated efficient only because these rare factor combinations are unique and impede unit comparison. As a result, an increased number of variables generate an increased number of efficient units (Jenkins and Anderson

2001). The more variables used for output generation, the higher the probability is that the observed unit has a unique (sub-) combination of factors without suitable benchmarks. In some DEA models, the share of efficient units is more than 50% of the observed DMUs (Dyson et al. 2001).

Ill-specified production function weights may result in implausible sales objectives. Ill-specified weights can render units efficient solely due to the ill-specification. Because efficient units are benchmarks for other DMU, the falsely identified benchmarks thwart the remaining benchmarking process.

To address the aforementioned DEA drawbacks, we suggest that the linear programming be replaced with econometric estimation of weights. The latter technique maximizes the overall model fit instead of individual relative efficiency. The multiple observations per unit provide information on the behavior of the production function over an interval of several periods instead of at a single point in time. This avoids ill-specification of individual production (response) weights. Because the individual weights maximize overall model fit instead of individual relative efficiency, having rare or extreme factor utilizations, outlier values, or a large number of variables no longer leads to increased probabilities of a unit being rated as efficient. In contrast to conventional DEA's linear programming weights, the econometrically estimated individual coefficients may be evaluated by significance information, and the response model fit can be evaluated by an R^2. Both help to identify misspecification and improve forecasting efficiency of benchmarking recommendations so that managers get more reliable benchmarking results.

We use multiple observations per unit so that the share of efficient units is considerably reduced, leading to an improved discrimination between units. Because the number of efficient units decreases, the share of inefficient units increases, leading to a larger potential for efficiency improvement.

In Section 4 we compare both models for a sales force benchmarking example according to the average individual sales volume that they can achieve with an orientation towards related individual benchmarking partners (i.e., sales objectives). The comparison of both model results suggests that a DEA with econometrically derived weights offers a much higher potential for improvement compared to the conventional DEA. In addition, we applied a holdout-validation, which demonstrates the conventional DEA's drawbacks, as given by implausible sales objectives and the very large share of efficient units.

3. Modeling relative efficiency using individualized econometrically estimated coefficients

3.1 Model framework

Because the application of our model takes place in a pharmaceutical sales force, we introduce our model in a sales force context. Nonetheless, the model is general enough to be applied to other benchmarking contexts. In the case of salesperson efficiency, the weights needed for a DEA are nothing more than individualized (i.e., salesperson-specific) parameter values of a sales response function (the weights reflect the conventional DEA feature 2 described in Section 2). We estimate idiosyncratic parameter values using the hierarchical Bayes model by Rossi, Allenby, and McCulloch (2005). With these weights, or parameter values, we can conduct all efficiency evaluations as recommended by DEA. In particular, we are able to calculate the ratio of the observed output to the estimated output as a function of the input for the observed unit, with the input factors being weighted by the individual coefficients. This gives us individual efficiency scores comparable to conventional DEA (the individual efficiency scores reflect the conventional DEA feature 1 described in Section 2). The units' efficiency scores using their own weights and factors are used for a relative ranking of DMUs according to relative efficiency. Other DMUs obtaining higher efficiency scores using the weights of the observed DMU can be used as benchmarks for the observed DMU. The benchmarking units generate higher output while using the same sales response parameter weights compared to the unit under consideration (the identification of superior benchmark units reflects the conventional DEA feature 3 described in Section 2). Possible output improvements result from the understanding that each observed unit should be able to generate as much output as each benchmarking DMU using the weights of the observed unit. Finally, the impact of the benchmarks is quantified by the level (size of the efficiency score) by which the benchmark unit outperforms the observed DMU. A benchmark unit whose efficiency score is far higher than the score of the observed DMU has a higher impact on the observed DMU than a benchmark unit whose score is only marginally higher than the efficiency of the observed DMU (the benchmarks' influence on the observed unit reflects the conventional DEA feature 4 described in Section 2). Replacing DEA's linear programming estimation based on just one data point per unit by econometric estimation based on multiple observations per unit, the advantages of both approaches can be combined. The main features of a conventional DEA are preserved while important improvements from the econometric framework are integrated.

3.2 Individualized sales response estimation

The first modeling step is to estimate the individual sales response function per salesperson. We use the hierarchical Bayes linear model described by Rossi, Allenby, and McCulloch (2005) for a semi-log sales response function, which is frequently used in sales response models (Doyle and Saunders 1990). The transformation of independent variables in natural logs allows for the use of linear estimation, which corresponds to the linear programming of conventional DEA. Furthermore, the semi-log form accounts for diminishing marginal effects so that there is no need to apply a DEA with decreasing returns-to-scale (Banker, Charnes, and Cooper 1984). The monthly sales volume is the dependent variable. On the independent side, there are usually variables that can be influenced by the salesperson, such as sales calls, and variables that cannot be influenced by the salesperson, such as indicators for customer potential (Skiera and Albers 1998). The estimated salesperson efficiency must focus on the controllable side of the sales response. We use spending on sales calls and marketing events per potential unit as controllable factors and territory-potential indicators as non-controllable variables. The customer base may be separated into several segments reflecting the differing sales potential. Finally, we account for carry-over effects from previous periods using stock variables. The response function is estimated at the salesperson level, leading to an individual set of coefficients for each salesperson unit. The actual response model design is usually dependent on the available data and can be found in the empirical Section 4 and in more detail in the Appendix Section A. Higher response coefficients characterize factors that have greater influence on the output generation process; lower coefficients reflect less influence. In contrast to conventional DEA, where individual relative efficiency is maximized, the econometrically estimated weights are based on multiple observations per unit and the maximization of response model fit. They can be assessed with respect to their significance, and a model R^2 can be calculated providing information on the goodness-of-fit.

3.3 Modeling relative efficiency and benchmarks

Based on the production function weights, DEA calculates efficiencies that can be used for benchmarking. Comparable to a conventional DEA, we define the efficiency of the observed unit by the output-input ratio in equation (1).

$$Efficiency_o = \frac{y_o}{\sum_{i=1}^{I} \beta_{i,o} x_{i,o}} \tag{1}$$

The $Efficiency_o$ of the currently observed unit o is given by the output y_o of unit o, divided by the aggregated weighted input of unit o. This input is given by the sum of all input factors $x_{i,o}$ over i (i=1,…,I) of unit o, each multiplied with the individual coefficient $\beta_{i,o}$ of input i of unit o. The denominator calculates a term similar to a conventional DEA model, which aggregates all input factors of the observed DMU to a single virtual input for further computation. $Efficiency_o$ measures the output-input ratio of the observed unit o using its own estimated weights and factors. As in a conventional DEA model, we are interested in the performance of other DMUs g (g=1,…,G), using the weights of unit o. If another unit g performs better than unit o using the weights of unit o, then this DMU g can be used as a benchmark for unit o because unit o should be as efficient as any other DMU using its own weights. Equation (2) describes this in formal notation. The ratio in (2) is comparable to the comparison of the observed unit to all other units in a conventional DEA model (Charnes, Cooper, and Rhodes 1978).

$$Efficiency_{g|o} = \frac{y_g}{\sum_{i=1}^{I} \beta_{i,o} x_{i,g}} \tag{2}$$

$Efficiency_{g|o}$ is the efficiency of unit g using the weights $\beta_{i,o}$ of unit o but using the input $x_{i,g}$ and output factors y_g of unit g. We compute $Efficiency_{g|o}$ for each DMU g. If $Efficiency_{g|o}$ is higher than $Efficiency_o$, unit g is a benchmark for unit o. The more $Efficiency_{g|o}$ exceeds $Efficiency_o$, the greater the influence benchmarking unit g has on unit o because the efficiency gap between g and o is larger. A smaller efficiency gap between $Efficiency_{g|o}$ and $Efficiency_o$ indicates a reduced impact of the benchmarking unit g on unit o, because benchmark unit g just slightly outperforms the observed unit o. If the observed unit o increases efficiency and closes the efficiency gap between $Efficiency_{g|o}$ and $Efficiency_o$, the efficiency of o increases, and the size of the efficiency gap decreases. This reflects a smaller impact of the benchmarking unit g that creates the efficiency gap between g and o. In comparison to conventional DEA scores, our measures of $Efficiency_o$ and $Efficiency_{g|o}$ are not limited to 100%.[1] For mutual comparability of relative efficiency scores across the whole sales force, we normalize the efficiency of each unit. Without normalization, efficiencies are only comparable between the observed DMU and its related benchmarks, but not comparable across the observed DMU, other observed DMUs, and their related benchmarks. Equation (3) describes the normalization in formal terms.

$$Efficiency_{g|o}^{normalized} = \frac{Efficiency_{g|o}}{Max\left(Efficiency_{h|o} \middle| h \in G\right)} \tag{3}$$

The normalized $Efficiency_{g|o}^{normalized}$ of g using the weights of o is computed by dividing the observed score $Efficiency_{g|o}$ by the maximum efficiency score over all h=1,…,G $Max(Efficiency_{h|o} \mid h \in G)$ using the weights of the observed unit o. The normalization of $Efficiency_o$ can be computed analogously by substituting $Efficiency_{g|o}$ for $Efficiency_o$ in equation (3).

We can visualize our results using a matrix that has as many rows and columns as observed units. For example, each row shows the observed unit o with its own weights and with the weights of each other unit g. The columns show one set of weights applied to each unit; each column contains the efficiency score of the observed unit o and the benchmarking units for unit o. The matrix diagonal from the upper left corner to the lower right corner contains the scores of the DMUs using their own weights. Table 1 shows an illustration using three DMUs.

[1] At this point, we neglect superefficiency models which generate efficiencies larger than 100% (Andersen and Petersen 1993) or effects which occur in the discussion of input-oriented versus output-oriented models and reciprocal efficiency scores (Cooper, Seiford, and Tone 2006).

	Column 1	Column 2	Column 3
Row 1	**DMU 1 using weight DMU 1**	DMU 1 using weight DMU 2	DMU 1 using weight DMU 3
Row 2	DMU 2 using weight DMU 1	**DMU 2 using weight DMU 2**	DMU 2 using weight DMU 3
Row 3	DMU 3 using weight DMU 1	DMU 3 using weight DMU 2	**DMU 3 using weight DMU 3**

Table 1: Benchmarking example.

The benchmarking in Table 1 takes place as follows. If the efficiency of *DMU 3 using weight DMU 1* is higher than the efficiency of *DMU 1 using weight DMU 1*, DMU 3 is a benchmark for DMU 1. If the efficiency of *DMU 2 using weight DMU 1* is also higher than the efficiency of *DMU 1 using weight DMU 1*, but smaller than *DMU 3 using weight DMU 1*, DMU 2 is also a benchmark for DMU 1, but the influence of DMU 3 on DMU 1 is higher. These influences correspond to the lambda values of the DEA envelopment form.

4. Empirical sales force application from the pharmaceutical industry

4.1 Data on a pharmaceutical sales force

We use panel data from a medium-sized German pharmaceutical company selling its products via a sales force in conjunction with additional marketing events (e.g., sponsored conferences). The firm has an extensive database that records all sales calls to and events for four physician segments (VIP, A, B and C), which were created according to their prescription potential. The company has access to IMS Health sales data and pharmacy coverage for 1,860 sales coverage units (SCU). There are 55 sales territories in Germany, each aligned to a single salesperson so that each territory consists of an average of 34 SCU. To comply with German privacy legislation, market research companies like IMS are not allowed to provide sales data on individual physicians and thus provide data only at an aggregate level for geographical units (SCUs). Each SCU represents an aggregation of at least six physicians so that no data can be traced back to the individual physician. The data contain multiple observations for each SCU for a period of 43 months between 2001 and 2004. The multiple data points allow for the econometric estimation of individualized (i.e., salesperson-specific) response parameters as discussed before. The first 36 months of the panel (2001 to 2003) are used for estimation purposes; the remaining seven months (2004) are used for holdout validation.

4.2 Individualized sales response estimation

As discussed in Section 3.2, we specify a semi-logarithmic sales response function. The transformation of independent variables into natural logs allows for the use of a linear estimation model that corresponds to the linear programming model of conventional DEA. Furthermore, the semi-log form accounts, by definition, for diminishing marginal effects so that it is not necessary to work with DEA models with decreasing-returns to scale (Banker, Charnes, and Cooper 1984). Because sales volume is frequently used as a performance indicator (Skiera and Albers 2008), we use the observed sales of each salesperson per SCU and per month as the dependent variable. On the independent variable side, we account for the factors that can be influenced by the salespersons, such as detailing spending (for sales calls) and event spending, and other potential factors that salespersons cannot influence. Another potential independent variable captures the SCU-specific pharmacy over- or under-coverage relative to the number of doctors in that SCU. Details of the specification of the response function can be found in the Appendix Section A. We use the average spending for each SCU per doctor, month, and segment, which is comparable to the call frequency (if not expressed in spending) and an important variable used for managing a sales force. We estimate territory- (salesperson) and segment-specific detailing-event coefficients, territory- (salesperson) specific pharmacy coverage coefficients, territory- (salesperson) specific intercepts, and error terms.

This specification helps us to account for two sources of heterogeneity. First, we account for observed heterogeneity within individual sales territories and across SCUs by having a SCU-specific potential factor that is multiplied by the individual detailing-event frequency variables and reflects SCU heterogeneity within the respective territory. Second, we account for unobserved heterogeneity across salespersons (territories) by estimating salesperson-specific individual response coefficients. The heterogeneity across SCUs (within territories) is mostly due to the number of doctors and their different distribution across segments, while individual salesperson coefficients (weights) reflect the salesperson-specific selling ability. The latter is required for the efficiency assessment. Table 2 displays point estimates and t-values for the first five example territories[2], the mean values across all 55 territories, the minimum and maximum values, and the standard deviation.

Var.	**Intercept**		**Detailing-Event Spending for Segment VIP**		**Detailing-Event Spending for Segment A**		**Detailing-Event Spending for Segment B**		**Detailing-Event Spending for Segment C**		**Pharmacy Coverage**	
DMU	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**
1	24875**	27.09	15.15**	26.82	15.16**	28.99	49.87**	18.19	47.67**	8.40	45.70**	15.13
2	13767**	29.63	21.38**	21.03	26.64**	29.16	19.24**	10.68	23.65**	5.85	72.28**	25.46
3	18532**	26.73	21.59**	35.28	16.68**	24.42	8.63**	4.03	43.51**	8.45	63.57**	17.61
4	18750**	17.71	13.43**	10.89	27.13**	24.70	12.50**	4.79	53.96**	10.80	40.97**	7.23
5	11697**	15.02	16.71**	17.67	22.88**	28.97	14.59**	7.30	94.54**	12.47	47.55**	13.72
Min	5168		10.96		7.43		-11.45		-127.16		27.44	
Max	29243		50.43		45.29		75.74		134.21		134.38	
Mean	**16368**		**21.28**		**25.64**		**20.87**		**32.42**		**70.50**	
SD	6459		6.85		7.52		18.35		53.04		24.29	
% not sig. at 5% level	0.00%		0.00%		3.64%		14.55%		21.82%		0.00%	
% not sig. at 10% level	0.00%		0.00%		0.00%		10.91%		14.55%		0.00%	

*significant at 10% level, **significant at 5% level. One-sided tests. Adjusted $R^2 = 70.6\%$

Table 2: Estimated coefficients and t-values.

The estimated salesperson level coefficients may be interpreted as econometrically derived DEA weights; for example, the first factor *Detailing-Event Spending per Doctor in Segment VIP* for DMU 1 has a weight of 15.15. The second factor of DMU 1, *Detailing-Event Spending per Doctor in Segment A,* has a comparable impact on the generated sales (dependent variable) with a weight of 15.16. In contrast to a conventional DEA, we can assess the significance of these weights. Both weights (the weight of the first mentioned input factor (15.15) and the weight of the second mentioned input factor (15.16)) are significant at the 5% level. Overall, the coefficients' mean values have a positive sign and plausible values, which is as expected. As reported in the row *Percentage not significant at 5% level,* 0.00% (3.64%,

[2] Here and in the following, the numbers of the territories have been arbitrarily renumbered so that results cannot be traced back to individuals. In addition, we only report results for some territories due to the same concern.

14.55%, 21.82%) of the coefficients are not significant at the 5% level for *Detailing-Event Spending per Doctor in Segment VIP (A*, *B, C),* and 0.00% (0.00%, 10.91%, 14.55%) of the coefficients are not significant at the 10% level for *Detailing-Event Spending per Doctor in Segment VIP (A, B, C)* as reported in the row *Percentage not significant at 10% level*. For all remaining variables, all coefficients are significant at least at the 10% level. This indicates a good stability for the estimated weights. The response model provides a very satisfactory fit with an adjusted R^2 of 70.6%. The conventional DEA does not offer similar robustness checks. Along with the problem that conventional linear programming weights can be misleading, the lack of significance levels and R^2 information is a serious drawback of the conventional model.

As described in Section 4.1, our data represent a panel of 43 monthly periods (2001 to 2004). The first 36 months (2001 to 2003) are used for the individualized response estimation. The remaining seven months in 2004 are used for a holdout validation of the estimated weights. Salesperson-specific response weights are used to predict each salesperson's sales volume in each of the remaining seven months of the year 2004. We evaluate the estimated coefficients by comparing the realized and predicted sales volumes in each of the seven holdout periods. Because sales volume data are at SCU level and the estimated weights are at salesperson level (territory level), we aggregate SCU level sales data on territory level by summing up the SCU level sales volumes for each territory. We calculate the Mean Absolute Percentage Error (MAPE) for each salesperson as the average of seven holdout months. Thus, monthly salesperson level data may be used for the evaluation. Formal details on the calculations can be found in the Appendix Section B. The average MAPE over all 55 salespersons is 15.10%. The smallest prediction error occurs for salesperson 29 (5.95%), and the largest occurs for salesperson 34 (24.75%); the standard deviation is 4.26%. Thus, the scale of the MAPE measures implies reasonable estimation quality. In Section 4.4 we compare these results to the prediction error of a conventional DEA.

4.3 Calculation of relative efficiencies and benchmarks

After the estimation and evaluation of factor weights, the weights are used to calculate relative efficiencies and benchmarks as described in equations (1), (2), and (3). Two minor adjustments of equation (1) and equation (2) are necessary. The adjustments affect the pharmacy coverage and the territory potential variables, which are not influenceable for the salespersons and should be adjusted accordingly. Mathematical details on the adjustments can be found in the Appendix Section D. Table 3, below, describes salesperson efficiencies for a sample of the first ten DMUs (salespersons) as an example.

Set of weights observed → **DMU factors used ↓**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	…
1	64%	55%	60%	60%	50%	**61%**	**50%**	**58%**	**61%**	62%	…
2	**75%**	71%	**72%**	74%	**66%**	**74%**	**64%**	**74%**	**74%**	77%	…
3	**71%**	62%	70%	68%	**62%**	**68%**	**63%**	**65%**	**64%**	67%	…
4	**84%**	**81%**	**82%**	84%	**74%**	**86%**	**81%**	**79%**	**74%**	82%	…
5	61%	60%	56%	59%	52%	**60%**	**48%**	**63%**	**61%**	61%	…
6	60%	56%	58%	58%	**59%**	56%	**49%**	**62%**	**61%**	60%	…
7	60%	54%	55%	57%	47%	**59%**	47%	**56%**	**59%**	60%	…
8	52%	46%	47%	49%	41%	49%	38%	50%	53%	51%	…
9	55%	48%	48%	51%	41%	52%	38%	**53%**	54%	52%	…
10	**88%**	**88%**	**83%**	**89%**	**79%**	**89%**	**72%**	**91%**	**100%**	100%	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…

Table 3: Examples of efficiencies for the first ten DMUs (salespersons).

The columns indicate from which unit the set of factor weights is taken. The rows describe the set of DMU factors used with the respective DMUs factor weights. As an example, the first cell in column one and row one (64%) describes DMU 1 with its own weights. The cell in column one, row two (75%) describes the output of DMU 2 divided by the inputs using the weights of DMU 1. The diagonal elements describe the relative efficiency ranking. For example, DMU 3 has an efficiency of 70%. The bold printed cells in each column are potential benchmarking units for the observed DMU because they are more efficient. Using the weights of DMU 3, DMU 4 has an efficiency of 82%. Therefore, DMU 4 is a benchmark for DMU 3. The efficiency score can be used as an indicator for which benchmarking unit has greater influence on a potential improvement of the observed unit. Taking observed DMU 3, the efficiency of benchmarking DMU 4 (82%) is greater than the efficiency of benchmarking DMU 2 (72%), but both efficiency scores are higher than the score of the observed DMU 3 on its own (70%). Thus, benchmarking

DMU 4 (82%) has greater influence than benchmarking DMU 2 (72%) on the improvement of observed DMU 3 (70%). These influences correspond to the conventional DEA's lambda values. Table 4 describes the top three benchmarks for the first ten DMUs.

Benchmarks & Efficiency (if there are more than three benchmarks for a DMU, only the top three benchmarks are reported)										
DMU	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
No. of Benchmarks	33	7	11	3	37	44	37	53	45	0
Top 3 Bench-mark DMUs	16(100%)	16(100%)	16(100%)	16(100%)	16(100%)	16(100%)	16(100%)	26(100%)	10(100%)	-
	10 (88%)	10 (88%)	10 (83%)	10 (89%)	29 (82%)	10 (89%)	4 (81%)	10 (91%)	16 (91%)	-
	25 (88%)	4 (81%)	25 (83%)	25 (84%)	10 (79%)	4 (86%)	10 (72%)	25 (86%)	25 (87%)	-

Table 4: Benchmarks for the first ten DMUs (salesperson).

The row labeled *No. of benchmarks* indicates the number of benchmarking units for each DMU. For example, DMU 1 has 33 benchmarks. In some cases, a large number of benchmarks exists (e.g., DMU 8 has 53 benchmarks, and DMU 9 has 45 benchmarks). This occurs if the observed DMU has a small relative efficiency (e.g., DMU 8 has an efficiency of only 50%, and DMU 9 has an efficiency of 54%). In these cases, many other units perform better than the observed DMU. Even if these units just slightly outperform the observed unit, they occur as a benchmark. This process generates a large number of benchmarks for the DMUs with a small relative efficiency. The benchmarking units and the related efficiencies are displayed in the row labeled *Top 3 Benchmark DMUs*. If there are more than three benchmarks for the observed unit, only those three benchmarks with the highest efficiencies are listed.

Benchmark units generate higher sales compared to the observed unit, using the observed unit's own weights and input factors of the benchmarking unit. Thus, a sales objective for each observed inefficient unit can be calculated. If the observed unit reaches this sales objective, it reaches the benchmark unit's efficiency level. Equation (4) describes the calculation of individual sales objectives in formal notation.

$$SO_o = y_o \cdot \sum_{b}^{B} \lambda_b \frac{Efficiency_{b|o}}{Efficiency_o} + \delta_o \cdot PhaCov_o \quad (4)$$

SO_o gives the sales objective for the observed unit *o*. y_o is the sales volume of the observed unit *o*.[3] λ_b is the influence of benchmark unit *b* (b=1,…,B) on the observed unit *o.* This influence is calculated as the efficiency gap between the observed unit's efficiency and the benchmark unit's efficiency, as described in Section 3.3. $Efficiency_o$ is the efficiency of the observed unit *o,* and $Efficiency_{b|o}$ is the efficiency of the observed unit's benchmark unit *b*. Thus, the ratio of $Efficiency_{b|o}$ to $Efficiency_o$ describes the sales volume improvement that the observed units can achieve with an orientation at benchmark unit *b*. Because we observe the top three benchmark units' *b*, the sales improvements are weighted by the individual benchmarks' influences λ_b. If the observed unit has less than three benchmarks, we use the maximum number of available benchmarks. Finally, we multiply the benchmark-weighted efficiency ratio by the sales volume y_o and add the pharmacy coverage effect $\delta_o \cdot PhaCov_o$, which cannot be influenced by the individual salesperson *o*. The ranking of the top five and last five salespersons, the observed salespersons' efficiencies, individual benchmarks, and related benchmark efficiencies are described in Table 5.[4]

Rank	**DMU**	**Efficiency** [in %]	**Sales (data)** [in €]	**Sales Objective** [in €]	**Improvement** [in %]	**Top 3 Benchmarks** (influence)
1	53	100.00%	69753	69753	0%	-
2	55	100.00%	51404	51404	0%	-
3	43	84.79%	54671	63753	16.61%	53 (100%), 55 (88%)
4	45	83.57%	38840	48241	24.20%	53 (100%), 55 (89%), 43 (84%)
5	10	80.17%	47048	57553	22.33%	53 (100%), 55 (95%), 43 (90%)
…						
51	34	48.47%	39569	70774	78.86%	55 (100%), 53 (94%), 43 (85%)
52	11	46.63%	33455	65770	96.59%	53 (100%), 45 (83%), 43 (81%)
53	3	46.57%	37208	71238	91.46%	53 (100%), 45 (81%), 55 (72%)
54	14	45.84%	36740	72860	98.31%	53 (100%), 55 (93%), 43 (85%)
55	17	42.19%	29761	60658	103.82%	53 (100%), 55 (90%), 45 (86%)
Mean		**65.50%**	**44908**	**64903**	**44.52%**	
Efficiency: Min.: 42.19%, Max.: 100.00%, Mean: 65.50%, SD: 12.47%						

Table 5: Sales force efficiency ranking of the recommended DEA with econometrically derived weights.

[3] Again, we need to adjust the sales volume of the observed unit *o* for the pharmacy coverage effects in equation (1) and (2). The pharmacy coverage adjustments are described in the Appendix Section E.

[4] Again, the territory numbers have been arbitrarily renumbered so that results cannot be traced back to individuals.

The column *Sales (data)* contains each salesperson's monthly sales volume (from the data and averaged across 36 monthly periods). The column *Sales Objective* contains the individual sales objectives, using the top three individual benchmark salespersons. The *Improvement* column indicates the individual percentage of sales improvement that can be achieved by an orientation towards the individual sales objective compared to the current average monthly sales volume, described in the column *Sales (data).* The last two lines contain sales force mean values and sales force efficiency distribution parameters.

The five top performing salespersons are DMUs 53, 55, 43, 45, and 10. The five worst performing salespersons are DMUs 34, 11, 3, 14, and 17. The average efficiency of the whole sales force is 65.50%. The minimum efficiency is 42.19%, the maximum efficiency is 100.00%, and the standard deviation is 12.47%. The benchmark's efficiency is indicated in brackets (e.g., unit 53 has an efficiency of 94% compared to DMU 34, which is 48.47% efficient using its own factors). The average salesperson's monthly sales volume is €44,908. If all salespersons successfully achieve the level of their benchmark units (i.e., the sales objective), the average monthly sales volume increases to €64,903 which represents an average improvement of 44.52% (€19,995) compared to the actual sales volume. Of course, such improvements are not likely to be fully realized in practice, but serve as an indication for potential gains. In the next section, we compare the individually achievable sales objectives and the individual holdout validation results with results based on a conventional DEA.

4.4 Comparison with a conventional DEA

We compute the results of a conventional DEA for comparison purposes. We use the basic CCR model proposed by Charnes, Cooper, and Rhodes (1978). The DEA is specified as an output-oriented model. We use the same factors as in the DEA model with econometrically derived weights. The sales volume is used as the output factor while the physician segment-specific "detailing-event frequency" and the pharmacy coverage are used as input factors. Because these variables are semi-log, decreasing returns to scale are implicitly taken into account. The input factors ("detailing and event frequency") are multiplied by the territory-specific prescription potential, as in the case of the DEA with econometrically derived weights (see equation (A1) in Appendix Section A). While calculating DEA weights, relative efficiency scores, and benchmark units, the pharmacy coverage is handled as a non-discretionary (non-influenceable) input variable. Because this calculation is done in a one-step process, a separated sales volume adjustment for the pharmacy coverage is not necessary when compared to the DEA with econometrically derived weights (see equation (C1) in Appendix Section C). The panel data are prepared for cross-sectional DEA by computing the salesperson-specific mean across the 36 monthly observations and summing the SCU-specific data at the sales territory level. Individual salespersons' sales objectives are calculated using the input factors of the top three benchmark units, which are determined by their influence (lambda value) on the observed inefficient DMU. For the calculation of sales objectives, the technique described in equation (4) can be used analogously for the DEA with econometrically derived weights. When salespersons achieve their individual sales objectives, they have become as efficient as their related benchmark unit. Table 6 describes the top five and bottom five ranked salespersons, related efficiencies, benchmark units, benchmark influences (lambda value), sales volumes (average across 36 monthly periods, given in the data), sales objectives, and the percentage sales improvement (if the sales objectives are achieved compared to the actual sales volume). The last two lines contain sales force mean values and efficiency distribution parameters.

Overall, there are 19 top-performing salespersons, which represent a share of 34.55%. The average efficiency is 87.86%, the minimum efficiency is 55.65%, and the standard deviation is 12.35%.

As in the DEA model with econometrically derived weights, we use the remaining seven months of the year 2004 for a holdout validation of the conventional DEA. Remember that we used the same panel data in both models, the only difference being that we used mean values across the 36 monthly periods and added up SCU data on a territory level in the conventional DEA model, reflecting the cross-sectional data requirements. Using salesperson-specific DEA coefficients and the remaining seven periods of 2004, we use the same holdout validation technique as applied to the DEA with econometrically derived weights.

The weights are used to predict each salesperson's sales volume in 2004. The MAPE between the realized and the predicted sales volume in 2004 evaluates the coefficients.

Rank	DMU	Efficiency [in %]	Sales (data) [in €]	Sales Objective [in €]	Improvement [in %]	Top 3 Benchmarks (influence)
1	53	100.00%	69753	-	-	-
2	55	100.00%	51404	-	-	-
3	43	100.00%	54671	-	-	-
4	45	100.00%	38840	-	-	-
5	20	100.00%	41817	-	-	-
…						
51	22	66.03%	38970	59450	52.55%	43 (0.04), 53 (0.22), 4 (0.77)
52	21	61.94%	37853	59368	56.84%	43 (0.58), 53 (0.35), 17 (0.17)
53	48	61.38%	35615	58088	63.10%	20 (0.16), 49 (0.26), 4 (0.68)
54	7	60.18%	39969	67076	67.82%	20 (0.27), 54 (0.19), 53 (0.58)
55	14	55.65%	36740	64550	75.70%	43 (0.01), 53 (0.45), 4 (0.62)
Mean		**87.86%**	**44908**	**51685**	**15.09%**	
Efficiency: Min.: 55.65%, Max.: 100,00%, Mean: 87.86, SD: 12.35%						

Table 6: Sales Force efficiency ranking of a conventional DEA.

Smaller values indicate a smaller gap between the realized and predicted sales volumes, implying higher estimation quality. Table 7 illustrates the validation results and compares the validation of both models.

Comparison	DEA with econometrically derived weights	Conventional DEA
Average MAPE (all 55 salespersons)	15.10%	23.93%
Minimum MAPE	5.95%	1.79%
Maximum MAPE	24.76%	76.36%
Standard Deviation	4.26%	15.18%

Table 7: Comparison of validation results between both DEA models.

In the conventional model, the average MAPE across all 55 salespersons is 23.93%. The smallest error is 1.79% (salesperson 34), the largest error is 76.36% (salesperson 49), and the standard deviation is 15.18%. A comparison between the validation of the DEA model with econometrically derived weights and the validation of the conventional DEA shows that the average error of the DEA with econometrically derived weights (15.10%) is 8.83% less compared to the conventional DEA (23.93%). The econometrically derived coefficients are more tightly focused (standard deviation: 4.26%) compared to the conventional DEA weights (standard deviation: 15.18%). A comparison of the maximum errors indicates that the conventional model tends to produce extreme values for the weights (maximum: 76.36%) compared to the econometrically derived weights (maximum: 24.76%).

In Section 1 and Section 2 we pointed out that the conventional DEA may lead to ill-specified weights. We assume ill-specification by the appearance of two different weight characteristics. First, weights are assumed to be ill-specified if they have extreme values compared to the respective sales force average (in standard deviations). Second, weights are also assumed to be ill-specified if they lead to benchmark units that create implausible sales objectives. This occurs if the sales objectives are smaller than the actual sales volume, erroneously implying that sales should optimally decline.

The holdout validation supports the assumption that, in the observed case, conventional DEA weights may be of inferior quality compared to the econometrically derived weights. As an example, we focus on the linear programming weights of salesperson 23 (fully efficient in the conventional DEA), who has extreme values compared to the average. The weight of the *Detailing-Event Spending per Doctor on Segment C* variable (0.0418) has a value 3.08 times that of the standard deviation (0.0121) from the sales force average (0.0045). Measured by the number of standard deviations, this weight has a value that varies extremely from most of the observed units, giving support to the assumption that this weight may be ill-specified.

To find evidence for the existence of the second effect, we calculate the sales objectives as described in equation (4), but instead use the benchmark units and benchmark influences (lambda values) identified by the conventional DEA model and the efficiency scores (for observed units and related benchmarks) based on the econometrically derived weights. Formal details on the modifications can be found in the Appendix Section E. A closer look at some individual salesperson's sales objectives makes the problem apparent. The whole sales force has 55 salespersons, of which 19 salespersons are fully efficient in the conventional DEA and do not have individual benchmark units. The remaining 36 salespersons do have individual benchmark units. If we apply the modifications in the sales objectives to those 36 salespersons that can be improved, 4 of those 36 salespersons (11.11%) create sales objectives that are smaller than the actual sales volume. Those sales objectives are implausible. They result from conventional DEA benchmark units that are identified based on ill-specified factor weights. Based on appropriately specified weights, as in our recommended DEA model with econometrically derived weights, these units would not have been identified as efficient benchmark units. Table 8 describes the implausible sales objectives and compares them to sales objectives based on the DEA with econometrically derived weights and to the actual sales volume.

Salesperson	**Sales (data)** [in €]	**Sales objective based on DEA with econometrically derived weights** [in €]	**Implausible sales objective using equation (E1)** [in €]	**Differences** [in €; in %]	
				Sales (data) vs. Implausible sales objective	DEA with econometrically derived weights vs. Implausible sales objective
Salesperson 5	43,062	50,431	36,476	-6,586 (-15.29%)	-13,955 (-27.67%)
Salesperson 30	42,333	53,831	35,461	-6,872 (-16.23%)	-18,370 (-34.13%)
Salesperson 7	40,802	53,481	38,482	-2,320 (-5.69%)	-14,999 (-28.05%)
Salesperson 10	47,048	57,553	41,305	-5,743 (-12.21%)	-16,248 (-28.23%)

Table 8: Ill-specification of conventional DEA weights leads to implausible sales objectives.

Salesperson 5 has a monthly sales volume of €43,062 (in the data). Its sales objective according to the DEA with econometrically derived weights is €50,431. Its sales objective using the benchmarks and benchmark influences of the conventional DEA, however, is €36,476, which is €6,586 (15.29%) less than the actual sales volume and €13,955 (27.67%) less than the recommended model with econometrically derived weights. Table 8 details three other example salespersons that have implausible sales objectives. These are the salespersons 30, 7, and 10. These implausible sales objectives give additional evidence to the assumption that the conventional DEA benchmark units may be identified based on ill-specified factor weights.

The very high share of efficient salespersons in the conventional model (34.55%) makes clear that the recommended DEA model with econometrically derived weights considerably reduces the share of efficient units. In this recommended model, only 3.64% (or 2 of the 55 salespersons) are fully efficient, while the average sales force efficiency is only 65.50% (compared to 87.86% in the conventional model). The ability to discriminate is highly improved, which is apparent based on the reduced number of efficient units. The reduced share of efficient units offers a higher opportunity for improvement because there are more units that can be improved. The 19 top performing salespersons of the conventional model do not have benchmarks because they are already rated as fully efficient. All other DMUs have benchmarks. In Table 6, a benchmark's impact is indicated in brackets (e.g., unit 53 is a benchmark for DMU 22 having an influence (lambda value) of 0.22). A salesperson's average actual monthly sales volume is €44,908. If all salespersons successfully achieve the level of their related benchmark units (i.e., the sales objective), the average monthly sales volume increases to €51,685, which is an average improvement of 15.09% (€6,777). The DEA with econometrically derived weights is 29.43% (= 44.52% - 15.09%) better than the conventional DEA model, which makes a difference in the sales volume of €13,178 (= €19,955 – €6,777) for each month and salesperson. If all salespersons successfully achieve their individual sales objectives, the DEA with econometrically derived weights allows for sales improvements of €13,170,300 (€19,955 for 55 salespersons in each of 12 months) while the conventional DEA allows for improvements of only €4,472,820. This creates a difference of €8,697,480 per year for the entire sales force. Table 9 summarizes the results for both models.

The combination of cross-sectional data and a large number of factors results in an additional drawback to the conventional model. If the observed unit faces unusual working conditions (e.g., a sales territory with a very low or very high selling potential), there may be a lack of other units that have similar working conditions and can be used as benchmarks. The large number of factors intensifies the problem because using a large number of factors makes it even harder to find comparable benchmarks with a similar factor configuration. Because the conventional DEA determines factor weights as much in favor of the observed units and maximizes the individual efficiency, the lack of suitable units for comparison renders the observed unit efficient. This process produces wrong benchmarks, solely due to the lack of data and the DEA efficiency maximization. The falsely identified benchmarks bias the remaining benchmarking effort by providing wrong orientation for other DMUs.

Comparison		DEA with econometrically derived weights	Conventional DEA	Difference (DEA with econ. derived weights vs. conventional DEA)
Share of salespersons being fully efficient (cannot be improved)	No. (%)	2 of 55 3.64%	19 of 55 34.55%	-17 -30.91%
Average sales force efficiency	(%)	65.50%	87.86%	-22.36%
Average salesperson's monthly sales improvement using sales objectives from the top three benchmarks	(€/month) (%)	€ 19,955 44.52%	€ 6,777 15.09%	€ 13,178 29.43%
Sales improvement for the whole sales force per year	(€/year)	€ 13,170,300	€ 4,472,820	€ 8,697,480

Table 9: Comparison of results for DEA with econometrically derived weights and conventional DEA.

An example from the current application is salesperson 17. His or her monthly sales volume is €29,761 while the average member of the sales force sells €44,908 per month. Thus, salesperson 17 has just 57.15% of the average sales volume, which would be qualified as a special working condition. As expected, salesperson 17 is rated efficient in the conventional DEA. In the DEA with econometrically derived weights, salesperson 17 is just 42.19% efficient. Thus, in management practice, the recommended model with econometrically derived weights gives additional credibility and robustness to the efficiency rating of each DMU. In the next section we point out the main managerial implications that can be derived from the comparison of both models.

5. Managerial implications

In management practice, the DEA with econometrically derived weights has some important advantages over a conventional model. A comparison of the holdout validation results in Section 4.4 demonstrates the smaller prediction error using the weights of the recommended DEA model compared to the conventional DEA's linear programming weights. The quality of the econometrically derived weights is supported by the ability to test their significance and the response model's goodness-of-fit via an R^2. Both are not provided in the conventional DEA. The implausible sales objectives of the conventional DEA, which are outlined in Section 4.4, support the assumption that the conventional weights are ill-specified and result in biased benchmarking results. Thus, it is doubtful that the conventional DEA derives appropriate benchmarking recommendations. In contrast, the predictive validity of the econometrically derived weights is much higher, leading to credible benchmarking recommendations.

In the DEA with econometrically derived weights, the multiple observations per unit and the maximization of model fit avoid the drawback of the conventional DEA (namely that a DMU that faces unusual working conditions may be rated efficient solely due to a lack of comparable units that face similar working conditions and which can be used as benchmarks). This effect, which is described in Section 4.4 in more detail, biases the remaining benchmarking processes because single benchmarks are identified falsely and erroneously orient the other units.

Along with the improved weights, the richer information (multiple observations per unit) used by the recommended DEA results in a smaller share of efficient units and a smaller average efficiency (see Section 4.4, Table 7 for details). This results in a higher share of units that can be improved and, thus, a higher potential for efficiency improvement vis-à-vis higher sales objectives. As depicted in Section 4.4, Table 7, the conventional DEA offers a monthly sales improvement of 15.09% per salesperson while the recommended DEA with econometrically derived weights offer improvements of 44.52% per salesperson per month. For the whole sales force, the aggregate improvement per year is €4,472,820 in the conventional DEA and €13,170,300 in the recommended model. This represents an improvement of €8,697,480 per year as a result of using the recommended model instead of the conventional DEA.

6. Conclusion and implications for further research

The recommended DEA with econometrically derived weights is a model that overcomes some drawbacks found in the conventional DEA. It is based on multiple observations per unit (panel data) while conventional DEA models are based on single observations per unit (cross-sectional data). The conventional model maximizes individual unit efficiency using linear programming while the econometrically derived weights maximize overall model fit. The conventional DEA's combination of single observations per unit, linear programming, and maximization of individual efficiencies can lead to ill-specified factor weights. The holdout validation in Section 4 demonstrates the improved predictive validity of the econometrically estimated DEA weights relative to linear programming coefficients. Although econometric estimation requires enhanced data, it offers improvements in addition to the improved coefficients. Because the efficiency of observed DMUs is no longer maximized individually and actually based on multiple observations per unit, the problem of units being rated efficient solely due to extreme factor utilizations can be avoided. In a conventional DEA, this effect can lead to a large share of efficient units, occasionally as high as 50% of the observed DMUs. The recommended model reduces the extreme factor utilizations and the number of efficient units considerably while increasing the ability to discriminate between units. Because there are less efficient units, the ability to improve efficiency (and thus increase output) increases.

Although the recommended model offers several advantages relative to the conventional DEA, it has some limitations. First, the individualized production response estimation requires panel data. In situations without multiple observations per unit, individualized unit level weights cannot be derived econometrically. Second, while conventional DEA is implemented in various software packages, the recommended model is not yet implemented in standard software and needs to be set up for each analysis. Concerning the model's advantages compared to a conventional DEA, further development of the DEA with econometrically derived weights may be a meaningful area for further research.

Appendix

A. Specification of an individualized sales response estimation

Equation (A1) describes the response function in formal terms. Because sales volume is frequently used as a performance indicator (Skiera and Albers 2008), we use the observed sales of each salesperson *g* per SCU *r* (r=1,…,R) and month *t* (t=1,…,T) as the dependent variable $Sales_{g,r,t}$.

$$\begin{aligned} Sales_{g,r,t} = \alpha_g + \sum_{s}^{S} \beta_{s,g} \cdot \left[\left(Pot_s \cdot Docs_{s,r} \right) \cdot \ln\left(\frac{Detailing_{s,r,t} + Events_{s,r,t}}{Docs_{s,r}} + 1 \right) \right] \\ + \delta_g \cdot PhaCov_r + \varepsilon_g \end{aligned} \tag{A1}$$

On the independent variable side, we account for the factors detailing spending (for sales calls) and event spending which can be influenced by the salespersons and non-influenceable potential factors. As Detailing and Event spending are highly collinear, we add up both spending variables and divide them by the number of physicians $Docs_{s,r}$ to make them comparable across SCUs. This provides the average spending for each SCU per doctor, month, and segment, which is comparable to the call frequency (if not expressed in spending) and is an important variable used for managing a sales force. We use stock variables of the total spending on detailing $Detailing_{s,r,t}$ plus events $Events_{s,r,t}$ in month t in the observed SCU r and segment s (s=1,…,S). For the stock variables, we use a monthly carryover of 0.9, which has proven to be the best fitting value in a pre-estimation stage. Accounting for non-influenceable potential factors, we multiply the detailing-event variable with SCU-specific prescription potential $Pot_s \cdot Docs_{s,r}$, containing the number of doctors $Docs_{s,r}$ in SCU r and segment s by the segment-specific potential factor Pot_s, which is based on firm data from previous periods. The second independent variable is an additional potential factor and includes the SCU-specific pharmacy over- or under-coverage $PhaCov_r$ relative to the number of doctors in that SCU. In equation (A1), α_g is the territory-specific intercept, and ε_g is the territory-specific error term. We estimate territory (salesperson) g- and segment s-specific detailing-event coefficients $\beta_{s,g}$, territory- (salesperson) specific pharmacy coverage coefficients δ_g, intercepts α_g, and error terms ε_g.

This specification helps to account for two sources of heterogeneity. First, it accounts for observed heterogeneity within individual sales territories and across SCUs. We account for this type of heterogeneity by having a SCU-specific potential factor ($Pot_s \cdot Docs_{s,r}$), which is multiplied by the individual detailing-event frequency variables and reflects SCU heterogeneity within the observed territory. Second, we account for unobserved heterogeneity between salespersons (territories) by estimating salesperson-specific individual response coefficients. The heterogeneity across SCUs (within territories) is mostly due to the number of doctors and their different distribution across segments while individual salesperson coefficients (weights) reflect the salesperson-specific selling ability. The latter is required for the efficiency assessment. The hierarchical Bayes linear model used for estimation is implemented with bayesm (Rossi, Allenby, and McCulloch 2005) using a Gibbs Sampler and a normal prior in R. We use 50,000 draws, discarding a burn-in of 5,000 draws.

B. Holdout validation of estimated sales response weights using MAPE

Our data represent a panel of 43 monthly periods (2001 to 2004). The first 36 months (2001 to 2003) are used for the individualized response estimation. The remaining seven months in 2004 are used for a holdout validation of the estimated weights. Salesperson-specific response weights are used to predict each salesperson's sales volume in each of the remaining seven months of 2004. We evaluate the estimated coefficients by comparing the realized and predicted sales volumes in each of the seven holdout periods. Because sales volume data are at the SCU level, and the estimated weights are at the salesperson level (territory level), we aggregate SCU level sales data to the territory level by summing up the SCU level sales volumes for each territory. We calculate the Mean Absolute Percentage Error (MAPE) for each salesperson as the average of seven holdout months. Thus, monthly salesperson level data may be used for the evaluation. Equation (B1) describes the MAPE in formal notation.

$$MAPE = \frac{1}{T}\sum_{t}^{T}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \cdot 100\% \tag{B1}$$

The MAPE calculates the average absolute deviation of the predicted and realized value as a percent of the realized value. t (t=1,…,T) is the currently observed validation period (we use seven monthly periods), y_t is the realized sales volume in t, and $\hat{y}_t$ is the predicted sales volume in t. The average MAPE over all 55 salespersons is 15.10%. The smallest prediction error occurs for salesperson 29 (5.95%), the largest error for salesperson 34 (24.75%), and the standard deviation is 4.26%. Thus, the scale of the MAPE measures implies reasonable estimation quality.

C. Adjusted equations (1) and (2) reflecting the pharmacy coverage and selling potential

A necessary adjustment in equations (1) and (2) affects the pharmacy coverage. Pharmacy coverage describes whether a certain SCU has more pharmacies than needed for the physicians in that SCU. It is a factor that cannot be influenced by a salesperson; hence, we subtract the pharmacy coverage influence from each individual salesperson's sales volume. The adjusted sales volume is used for the calculation of individual relative efficiencies and benchmark units. Equation (C1) describes this adjustment in formal notation.

$$y_g^{no\_pha\mathrm{cov}} = y_g - \delta_g \cdot PhaCov_g \qquad \text{(C1)}$$

$PhaCov_g$ denotes the pharmacy coverage in territory g. δ_g is the salesperson g-specific pharmacy coverage coefficient as estimated in Section 4.2 (remember that each territory g is covered by a single salesperson g). y_g is the salesperson g-specific sales volume, including the pharmacy coverage effect (as described in the data). $y_g^{no\_pha\mathrm{cov}}$ is the salesperson g-specific sales volume without pharmacy coverage effect. Equation (C2) is a modification of equation (2) using the pharmacy coverage adjusted sales volume $y_g^{no\_pha\mathrm{cov}}$ in the numerator.

A second adjustment in equation (C2) affects the territory potential. As described in equation (2), we are interested in the output (sales volume), which is generated by all other units g (salespersons) using the weights of the observed unit o. In other words, when using the factors of all other units g in the territory of the observed unit o (i.e., using weights of the observed unit o), we need to account for the observed unit's selling potential. We modify equation (2) so that we do not solely use the weights of the observed unit o for all factors i, $\beta_{i,o}$ but also use the selling potential of the observed unit o and physician segment s, $Pot_s \cdot Docs_{s,o}$. Thus, for each territory g and input factor i, $x_{i,g}$, is split into the "detailing-event frequency" for input i, $Freq_{i,g}$, and the segment s- and territory g-specific prescription potential $Pot_s \cdot Docs_{s,g}$, so that $x_{i,g} = Pot_s \cdot Docs_{s,g} \cdot Freq_{i,g}$. The "detailing-event frequency", $Freq_{i,g}$, contains the detailing and event spending per doctor for input i in territory g, as described in Section 4.2. Both modifications of equation (2) are described in equation (C2).

$$Efficiency_{g|o} = \frac{y_g^{no\_pha\,\mathrm{cov}}}{\sum_{i=1}^{I} \beta_{i,o} \cdot Pot_s \cdot Docs_{s,o} \cdot Freq_{i,g}} \tag{C2}$$

The denominator reflects the input factors of salesperson *g*, $Freq_{i,g}$ weighted with the segment- and territory-specific selling potential of salesperson *o*, $Pot_s \cdot Docs_{s,o}$, and the regression coefficients for observed unit *o*, $\beta_{i,o}$, The numerator includes the pharmacy coverage adjusted sales volume $y_g^{no-pha\,\mathrm{cov}}$.

D. Pharmacy coverage correction of sales objectives

The pharmacy coverage is a factor that cannot be influenced by the observed salesperson; hence, we subtract the pharmacy coverage influence from each salesperson's sales volume as described in Appendix Section C. For the calculation of the sales objective of unit *o*, SO_o, we use the pharmacy coverage adjusted sales volume, $y_o^{no-pha\,\mathrm{cov}}$, instead of the non-adjusted sales volume, y_o. Equation (D1) describes the adjustments of equation (4) in formal notation.

$$SO_o = y_o^{no\_pha\,\mathrm{cov}} \cdot \sum_{b}^{B} \lambda_b \frac{Efficiency_{b|o}}{Efficiency_o} + \delta_o \cdot PhaCov_o \tag{D1}$$

The interpretation of the variables aside from $y_o^{no-pha\,\mathrm{cov}}$ is the same as described in Section 4.3.

E. Sales objectives using conventional DEA benchmarks and influences

We pointed out that the conventional DEA may lead to ill-specified weights. We assume weights to be ill-specified if they lead to benchmark units that create implausible sales objectives. This occurs if the sales objectives are smaller than the actual sales volume, thus implying that a sales decline is optimal. To find evidence for the existence of this effect, we calculate the sales objectives as described in equation (4) but use the benchmark units and benchmark influences (lambda values) identified by the conventional DEA model and the efficiency scores (for observed units and related benchmarks) based on the econometrically derived weights. For this purpose, we use a modification of equation (4), which is described in equation (E1).

$$SO_o^{ill-spec.} = y_o^{no\_pha\mathrm{cov}} \cdot \sum_{b^{conv.}}^{B^{conv.}} \lambda_b^{conv.} \frac{Efficiency_{b|o}^{stat.}}{Efficiency_o^{stat.}} + \delta_o^{stat.} \cdot PhaCov_o \tag{E1}$$

The efficiency scores $Efficiency_{b|o}^{stat.}$ and $Efficiency_o^{stat.}$, the pharmacy coverage variable $PhaCov_o$, the related weight $\delta_o^{stat.}$, and the sales volume $y_o^{no\_pha\mathrm{cov}}$, remain as described in equation (4) and equation (D1), respectively, and as used in the DEA model with econometrically derived weights. However, in contrast to the aforementioned model, we use the top three benchmark units identified by the conventional DEA $b^{conv.}$ ($b^{conv.} = 1,\ldots,B^{conv.}$) and their related influences, $\lambda_b^{conv.}$. The resulting sales objective for the observed unit o, $SO_o^{ill-spec.}$, provides additional evidence of a possible ill-specification of conventional DEA weights by creating implausible sales objectives.

References

Aly, H. Y., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation. *The Review of Economics and Statistics*, 72 (2), 211-218.

Andersen, P. and Petersen, N. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science*, 39 (10), 1261-1264.

Avkiran, N. K. (1999): An application reference for data envelopment analysis in branch banking: helping the novice researcher, *International Journal of Bank Marketing*, 17 (5), 206-220.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technological and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30 (9), 1078-1092.

Boles, J. S., Donthu, N., Lohtia, R. (1995). Salesperson Evaluation Using Relative Performance Efficiency: The Application of Data Envelopment Analysis. *Journal of Personal Selling and Sales Management*, 15 (3), 31-49.

Charnes, A. A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6), 429-444.

Charnes, A. A., Clark, C. T., Cooper, W. W., and Golany, B. (1985a). A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Forces. *Annals of Operations Research*, 2, 95–112.

Charnes, A. A., Cooper, W. W., Lewin, A. Y., Morey, R. C., and Rousseau, J. (1985b). Sensitivity and Stability Analysis in DEA. *Annals of Operations Research*, (2), 139-156.

Charnes, A. A., Cooper, W. W., Lewin, A. Y., and Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer: Boston.

Cooper, W. W., Seiford, L. M., and Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses. With DEA-Solver Software and References*. Springer: New York.

Donthu, N., Hershberger, E. K., and Osmonbekov, T. (2005). Marketing productivity using data envelopment analysis. *Journal of Business Research*, 58 (11), 1474-1482.

Doyle, P. and Saunders, J. (1990). Multiproduct Advertising Budgeting. *Marketing Science*, 9 (2), 97-113.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132 (2), 245-259.

Ferrier, G. D. and Lovell, C. A. K. (1990). Measuring Cost Efficiency in Banking. Econometric and Linear Programming Evidence. *Journal of Econometrics*, 46 (1-2), 229-245.

Giokas, D. and Vassiloglou, M. (1991). A goal programming model for bank assets and liabilities management, *European Journal of Operational Research*, 50 (1), 48-60.

Jenkins, L. and Andersen, M. (2001). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, 147 (1), 51-61.

Kamakura, W. A., Lenartowicz, T., and Ratchford, B. (1996). Productivity Assessment of Multiple Retail Outlets. *Journal of Retailing*, 72 (4), 333-356.

Mahajan, J. (1991). A data envelopment analytic model for assessing the relative efficiency of the selling function. *European Journal of Operational Research*, 53 (2), 189-205.

Oral, M. and Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches. *European Journal of Operational Research*, 46 (3), 282-294.

Rhonda, R. T., Barr, R. S., Cron, W. L.; Slocum Jr., J. W. (1998). A process for evaluating retail store efficiency: a restricted DEA approach. *International Journal of Research in Marketing*, 15 (5) 487-503.

Rossi, P. E. and Allenby, G. (2003). Bayesian statistics and marketing, *Marketing Science*, 22 (3), 304-328.

Rossi, P. E., Allenby, G., and McCulloch, R. (2005). *Bayesian statistics and marketing*. Wiley: Chichester.

Sherman, D. H. and Gold, F. (1985). Bank Branch Operating Efficiency. Evaluation with Data Envelopment Analysis. *Journal of Bank and Finance*, 9 (2), 297-315.

Skiera, B. and Albers, S. (1998). COSTA: Contribution Optimizing Sales Territory Alignment. *Marketing Science*, 17 (3), 196-213.

Skiera, B. and Albers, S. (2008). Prioritizing Salesforce Decision Areas for Productivity Improvements Using a Core Sales Response Function. *Journal of Personal Selling and Sales Management*, 28 (2), 145-154.

Zhu, J. (2003). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Kluwer: Boston.