MIS Quarterly

RESEARCH NOTE

A DATA ENVELOPMENT ANALYSIS APPROACH TO ESTIMATE IT-ENABLED PRODUCTION CAPABILITY

Sezgin Ayabakan
University of Baltimore, 1420 N. Charles Street,
Baltimore, MD 21201 U.S.A. {sayabakan@ubalt.edu}

Indranil R. Bardhan and Zhiqiang (Eric) Zheng
University of Texas at Dallas, 800 W. Campbell Road, SM 3,
Richardson, TX 75080 U.S.A. {bardhan@utdallas.edu} {ericz@utdallas.edu}

Appendix A

Correlation Matrix and Estimation Results

Table A1. Correlation Matrix

	Variables	v1	v2	v3	v4	v5	v6	v7	v8
v1	Margin	1							
v2	ProdCAP	0.25*	1						
v3	ITUsage	0.16*	0.06	1					
v4	Capex	0.11	0.01	0.06	1				
v5	TrainCost	0.07	0.05	0.04	0.13*	1			
v6	Size	-0.08	-0.08	0.41*	-0.12*	-0.03	1		
v7	Age	-0.01	-0.12*	0.07	-0.01	0.05	0.20*	1	
v8	PlantType	-0.01	0.05	-0.20*	0.01	-0.02	-0.30*	-0.02	1

*Statistically significant at p = 0.05

Table A2. Estimation Results for Production Capability Model (Without ITSpend)						
System	System 1		System 2		System 3	
Model	Model 1	Model 2	Model 1	Model 3	Model 1	Model 4
Dependent Variable	ProdCap	Margin	ProdCap	Margin	ProdCap	Margin
Intercept	1.095 (1.123)	36.024*** (9.014)	1.296 (1.152)	48.31*** (9.468)	1.095 (1.123)	41.07*** (9.137)
ProdCap	–	10.617*** (2.283)	–	–	–	9.813*** (2.281)
ITUsage	0.278** (0.15)	–	0.323** (0.155)	3.895*** (1.241)	0.278** (0.15)	3.03*** (1.194)
Capex	-0.02 (0.032)	0.57** (0.26)	-0.022 (0.032)	0.464** (0.269)	-0.02 (0.032)	0.49** (0.259)
TrainCost	1.208* (0.743)	0.245 (5.604)	1.381** (0.779)	3.31 (5.641)	1.208* (0.743)	0.026 (5.545)
Size	-0.119 (0.149)	-1.6* (1.101)	-0.163 (0.153)	-2.966*** (1.234)	-0.119 (0.149)	-2.771*** (1.183)
Age	-0.069** (0.035)	0.214 (0.277)	-0.076** (0.035)	0.019 (0.293)	-0.069** (0.035)	0.222 (0.274)
PlantType	0.069 (0.329)	-1.864 (2.608)	0.107 (0.339)	-0.947 (2.736)	0.069 (0.329)	-1.357 (2.588)
F-Val	2.9515	3.4579	3.0056	2.4878	2.9514	3.7841
R^2	0.1335	0.1529	0.1356	0.1150	0.1335	0.1760
Adj R^2	0.0919	0.1123	0.0941	0.0725	0.0919	0.1330
Heteroscdasticity Adjustment	No	Yes	No	No	No	Yes

Industry dummies are included in all estimation models. Significant one-sided * at $p < 0.10$; ** at $p < 0.05$; and *** at $p < 0.01$. Standard errors are shown in parentheses. Sobel Mediation test p = 0.03 and Goodman Mediation test p = 0.025 (one-sided p-values).

Appendix B

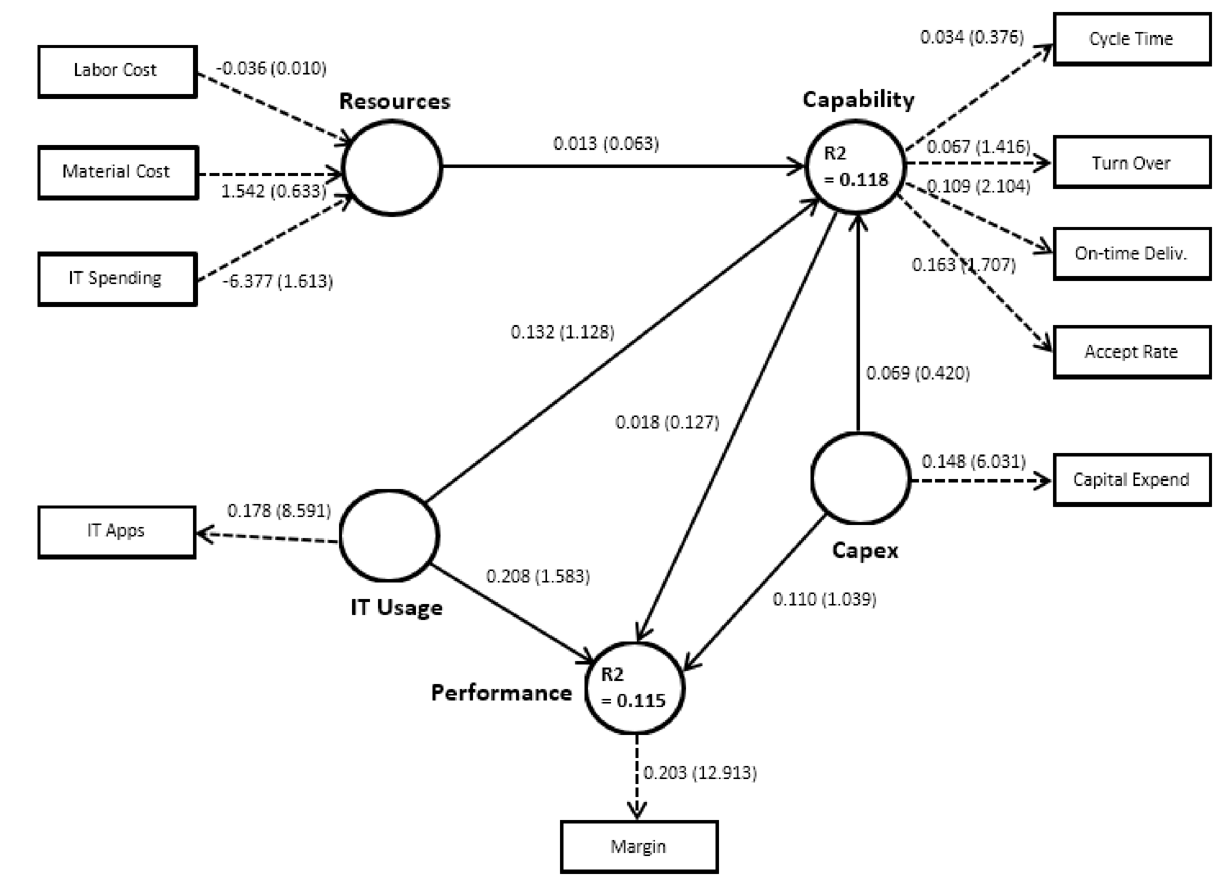
Comparison of DEA-Based Methods with Other Approaches

We compared our DEA-based approach for conceptualizing production capability with three other approaches commonly used in the RBV literature: (1) structural equation modeling (SEM); (2) stochastic frontier estimation (SFE); and (3) principal component analysis (PCA). We compared the R^2 and Adj R^2 values of these estimation methods with those obtained from the DEA approach, as reported in Table 7. Based on the greater R^2 and Adj R^2 values of the DEA-based methods, we concluded that our DEA approach exhibited greater explanatory power in explaining variations in plant performance.

Next, we provide further details of our estimation for each of these alternative methods.

Structural Equation Modeling

One stream of the literature on capabilities conceptualizes and operationalizes capabilities as a latent variable governing manifested measurement items using the SEM techniques (Schroder et al. 2002). Typically these studies involve survey questions that are designed to elicit responses (measurement items) based on perceptions about competencies and capabilities associated with different functional areas (Bharadwaj et al. 2007; Pavlou and El Sawy 2010).



** PLS model controls for Training Cost, Age, Size, Plant Type and Industry Dummies

Figure B1. PLS Estimation Results

In order to make a direct comparison between DEA- and SEM-based methods, we created two constructs, each representing a latent variable that governs inputs and outputs to our DEA model respectively. The construct "Resources" captures *Labor*, *Material Costs*, and *IT Spending* in a formative manner; and the construct "Capability" incorporates *CycleTime*, *TurnRate*, *OnTime*, and *AcceptRate* in a reflective manner.

We then used partial least squares (PLS) techniques to estimate the SEM model specified in Figure B1. PLS was preferred here to LISREL because of the presence of the "Heywood cases," in which some of the loadings can be negative (Fornell and Bookstein 1982). We used SmartPLS 2.0 to estimate the path coefficients as well as error variances. Figure B1 depicts the estimated path coefficients, as well as the t-values obtained from bootstrapping.

According to Chin (1998), existing goodness of fit measures assume that all measures in the assumed model are reflective and are related to how strongly the model accommodates sample covariances. However, some SEM procedures, such as PLS, have different objective functions and allow for formative measures. It is suggested that more attention should be paid to the fit of the SEM model when both reflective and formative constructs are present. In addition, Bollen and Long (1993) also suggest that fit of the components of a model, specifically R^2, can provide insight into the choice of a goodness-of-fit index. For this reason, we focused on R^2 to evaluate the fit of the SEM model. We observed that the R^2 of Model 1 in the DEA-based approach was 0.130, whereas in SEM it was 0.118. Similarly, the R^2 of Model 4 using our approach was 0.181, whereas it dropped to 0.115 using SEM. Of greater importance, none of the path coefficients in the SEM model were significant except *ITUsage* ➔ *Margin* (one-sided p-value = 0.058), rendering an overall, insignificant model. We also noted that the latent variable conceptualization of capability did not capture the relative capability across plants.

Stochastic Frontier Estimation (SFE)

The key limitation of SFE compared with DEA is that the former only accommodates a single output, while it is common for firms to make tradeoffs between multiple outputs. Nevertheless, for the purpose of comparing SFE with the DEA approach in our case, we factorized the multiple outputs into a single factor, *Fact_Out*, using principal component analysis (PCA), while maintaining the same input set. We followed the same procedure specified in Li et al. (2010) in developing the production function. We estimated the technical efficiency scores using a half-normal distribution for the inefficiency variable in the SFE model (Battese and Coelli 1988). We then applied "systems of equations" estimation using the SFE-based efficiency scores. These results are presented in Table B1. We observed that the R^2 of Model 1 of the SFE approach was lower than the corresponding values in the DEA approach (i.e., 0.113 versus 0.130). Likewise, the R^2 of Model 4 decreased from 0.181 to 0.119 when SFE was applied.

Table B1. Estimation Results with Stochastic Frontier Estimation

System	System 1		System 2		System 3	
Model	**Model 1**	**Model 2**	**Model 1**	**Model 3**	**Model 1**	**Model 4**
Dependent Variable	**Fact_Out**	**Margin**	**Fact_Out**	**Margin**	**Fact_Out**	**Margin**
Intercept	0.616*** (0.082)	36.238*** (10.349)	0.616*** (0.082)	48.31*** (9.468)	0.616*** (0.082)	43.67*** (10.49)
Fact_Out	–	10.341* (7.382)	–	–	–	7.527 (7.33)
ITUsage	0.022** (0.011)	–	0.022** (0.011)	3.895*** (1.241)	0.022** (0.011)	3.73*** (1.251)
Capex	0.001 (0.002)	0.556** (0.271)	0.001 (0.002)	0.464** (0.269)	0.001 (0.002)	0.457** (0.269)
TrainCost	0.04 (0.049)	3.478 (5.734)	0.04 (0.049)	3.31 (5.641)	0.04 (0.049)	3.012 (5.648)
Size	0 (0.011)	-1.563* (1.158)	0 (0.011)	-2.966*** (1.234)	0 (0.011)	-2.967*** (1.234)
Age	-0.004* (0.003)	0.05 (0.299)	-0.004* (0.003)	0.019 (0.293)	-0.004* (0.003)	0.049 (0.294)
PlantType	0.057*** (0.024)	-2.198 (2.797)	0.057*** (0.024)	-0.947 (2.736)	0.057*** (0.024)	-1.373 (2.767)
F-Value	2.43	1.83	2.43	2.488	2.43	2.386
R^2	0.113	0.087	0.113	0.115	0.113	0.119
Adj R^2	0.070	0.043	0.070	0.073	0.070	0.073
N	263		263		263	

Industry dummies are included in all estimation models. Significant one-sided at $p < 0.10$; ** at $p < 0.05$; and *** at $p < 0.01$. Standard errors are shown in parentheses.

Principal Component Analysis (PCA)

PCA has been used to conceptualize capability as a driver of performance. This is typically done by combining multiple process output measures into a single construct using the loadings derived from PCA as the weights (e.g., Ray et al. 2005; Rosenzweig et al. 2003). However, there may be various dimensions of outcomes in analyzing the operational and financial performance of organizations (Venkatraman and Ramanujam 1986). For example, indicators for operational performance may include innovation and productivity, while financial performance indicators may include earnings growth and stock price. Often, these disparate dimensions of outcomes do not converge (Combs et al. 2005). Therefore, one of the challenges of merging multiple outputs into a single construct lies in the possible tradeoffs among these various performance measures.

We applied PCA to our output variables and transformed them into one factor, in order to check if a single construct of output performance can satisfactorily represent plant production capability. We used PCA to merge *CycleTime*, *TurnRate*, *OnTime*, and *AcceptRate* into one factor, *Factor_Out*. We present the results obtained from the system of equations estimation using this derived factor in Table B2. We observed that *Factor_Out* failed to explain the variations in *Margin* in Model 2 as well as Model 4. In Model 1, none of the input variables appeared to be significant determinants of *Factor_Out* and most of the control variables were insignificant. In terms of R^2, our DEA-based approach exhibited better fit across all models.

Table B2. Estimation Results with Principal Component Analysis

System	System 1		System 2		System 3	
Model	Model 1	Model 2	Model 1	Model 3	Model 1	Model 4
Dependent Variable	Factor_Out	Margin	Factor_Out	Margin	Factor_Out	Margin
Intercept	0.173 (0.533)	42.341*** (9.428)	0.173 (0.533)	47.933*** (9.404)	0.173 (0.533)	47.896*** (9.421)
Factor_Out	–	0.75 (1.182)	–	–	–	0.348 (1.171)
ITUsage	0.141** (0.07)	–	0.141** (0.07)	3.755*** (1.257)	0.141** (0.07)	3.717*** (1.268)
Capex	0.016 (0.015)	0.528** (0.273)	0.016 (0.015)	0.439* (0.269)	0.016 (0.015)	0.436* (0.27)
TrainCost	0.244 (0.309)	3.277 (5.892)	0.244 (0.309)	2.963 (5.758)	0.244 (0.309)	2.873 (5.777)
LaborCost	0.004 (0.007)	–	0.004 (0.007)	–	0.004 (0.007)	–
MaterialCost	-0.002 (0.004)	–	-0.002 (0.004)	–	-0.002 (0.004)	–
ITSpend	-0.014 (0.023)	–	-0.014 (0.023)	–	-0.014 (0.023)	–
Size	-0.022 (0.067)	-1.568* (1.167)	-0.022 (0.067)	-2.975*** (1.231)	-0.022 (0.067)	-2.971*** (1.234)
Age	-0.023* (0.016)	0.053 (0.296)	-0.023* (0.016)	0.058 (0.29)	-0.023* (0.016)	0.064 (0.291)
PlantType	0.317** (0.149)	-1.826 (2.781)	0.317** (0.149)	-0.929 (2.716)	0.317** (0.149)	-1.025 (2.743)
F-Val	2.192	1.705	2.192	2.492	2.192	2.311
R^2	0.124	0.081	0.124	0.114	0.124	0.115
Adj R^2	0.071	0.037	0.071	0.072	0.071	0.069
Heteroscedasticity	No	Yes	No	Yes	No	Yes

Industry dummies are included in all estimation models. Standard errors are shown in parentheses. Significant one-sided * at $p < 0.10$; ** at $p < 0.05$; and *** at $p < 0.01$.

Appendix C

Robustness Check on DEA Sample Size

The DEA literature suggests using large samples for DEA calculation in order to obtain statistical validity in two-stage estimations, where DEA estimation is followed by a regression analysis (Banker 1993; Iyer et al. 2013). For this reason, as a robustness check of the sensitivity of our results to sample size, we excluded industries with less than 30 observations. Hence, we only kept the industries of *Chemicals*, *Metals*, *Machinery*, and *Electrical*. The total number of observations in our sample decreased to 209, with the exclusion of *Nondurables*, *Transportation*, and *Miscellaneous* industries. Our regression results of system of equations estimation are reported in the Table C1. Accordingly, our results were consistent with this additional analysis.

Table C1. System of Equations Estimation Results

System	System 1		System 2		System 3	
Model	**Model 1**	**Model 2**	**Model 1**	**Model 3**	**Model 1**	**Model 4**
Dependent Variable	**ProdCap**	**Margin**	**ProdCap**	**Margin**	**ProdCap**	**Margin**
Intercept	1.67* (1.295)	32.032*** (10.068)	1.67* (1.295)	49.865*** (10.762)	1.67* (1.295)	38.123*** (10.244)
ProdCap	–	13.609*** (2.411)	–	–	–	12.951*** (2.396)
ITUsage	0.233* (0.168)	–	0.233* (0.168)	3.99*** (1.395)	0.233* (0.168)	3.166*** (1.326)
Capex	-0.035 (0.038)	0.783*** (0.301)	-0.035 (0.038)	0.538** (0.32)	-0.035 (0.038)	0.647** (0.303)
TrainCost	0.573 (0.713)	3.36 (5.689)	0.573 (0.713)	4.887 (5.917)	0.573 (0.713)	3.041 (5.607)
Size	-0.314** (0.176)	-0.745 (1.248)	-0.314** (0.176)	-3.061** (1.421)	-0.314** (0.176)	-2.052* (1.337)
Age	-0.042 (0.039)	0.159 (0.312)	-0.042 (0.039)	0.047 (0.336)	-0.042 (0.039)	0.161 (0.309)
PlantType	0.218 (0.369)	-3.442 (2.885)	0.218 (0.369)	-2.335 (3.073)	0.218 (0.369)	-2.955 (2.86)
F-Val	1.613	4.421	1.613	1.907	1.613	4.67
R^2	0.075	0.182	0.075	0.087	0.075	0.206
Adj R^2	0.033	0.145	0.033	0.046	0.033	0.166
Heteroscedasticity Adjustment	No	Yes	No	No	No	Yes

Industry dummies are included in all estimation models. Standard errors are shown in parentheses. Significant one-sided * at $p < 0.10$; ** at $p < 0.05$; and *** at $p < 0.01$. Sobel Mediation test $p = 0.09$ and Goodman Mediation test $p = 0.08$ (one-sided p-values).

References

Banker, R. D. 1993. "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation," *Management Science* (39:10), pp. 1265-1273.

Battese, G., and Coelli, T. 1988. "Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data," *Journal of Econometrics* (38:3), 387-399.

Bharadwaj, S., Bharadwaj, A., and Bendoly, E. 2007. "The Performance Effects of Complementarities between Information Systems, Marketing, Manufacturing, and Supply Chain Processes," *Information Systems Research* (18:4), pp. 437-453.

Bollen, K. A., and Long, J. S. 1993. *Testing Structural Equation Models*, Newbury Park, CA: Sage.

Chin, W. W. 1998. "Issues and Opinion on Structural Equation Modeling," *MIS Quarterly* (22:1), pp. vii-xvi.

Combs, J. G., Crook, T. R., and Shook, C. L. 2005. "The Dimensionality of Organizational Performance and its Implications for Strategic Management Research," *Research Methodology in Strategy snd Management* (2), pp. 259-286.

Fornell, C., and Bookstein, F. L. 1982. "Two Structural Equation Models : LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research* (19:4), pp. 440-452.

Iyer, A., Saranga, H., and Seshadri, S. 2013. "Effect of Quality Management Systems and Total Quality Management on Productivity Before and After: Empirical Evidence from the Indian Auto Component Industry," *Production and Operations Management*, (22:2), pp. 283-301.

Li, S., Shang, J., and Slaughter, S. A. 2010. "Why Do Software Firms Fail? Capabilities, Competitive Actions, and Firm Survival in the Software Industry from 1995 to 2007," *Information Systems Research* (21:3), pp. 631-654.

Pavlou, P., and El Sawy, O. 2010. "The Third Hand: IT-Enabled Competitive Advantage in Turbulence through Improvisational Capabilities," *Information Systems Research* (21:3), pp. 443-471.

Ray, G., Muhanna, W. A., and Barney, J. B. 2005. "Information Technology and the Performance of the Customer Service Process: A Resource-Based Analysis," *MIS Quarterly* (29:4), pp. 625-652.

Rosenzweig, E. D., Roth, A. V., and Dean Jr., J. W. 2003. "The Influence of an Integration Strategy on Competitive Capabilities and Business Performance: An Exploratory Study of Consumer Products Manufacturers," *Journal of Operations Management* (21:4), pp. 437-456.

Schroeder, R. G., Bates, K. A., and Junttila, M. A. 2002. "A Resource Based View of Manufacturing Strategy and the Relationship to Manufacturing Performance," *Strategic Management Journal* (23:2), pp. 105-117.

Venkatraman, N., and Ramanujam, V. 1986. "Measurement of Business Performance in Strategy Research: A Comparison of Approaches," *Academy of Management Review* (11:4), pp. 801-814.