

Determinants of Commercial Banks' Efficiency: Evidence from 11 CEE Countries*

Dana PANČUROVÁ—University of Economics in Bratislava, Faculty of Business Economics in Košice, Slovakia (dana.pancurova@gmail.com)

Štefan LYÓCSA—University of Economics in Bratislava, Faculty of Business Economics in Košice, Slovakia (stefan.lyocsa@gmail.com), *corresponding author*

Abstract

This study estimates bank efficiencies and their determinants for a sample of 11 Central and Eastern European Countries (CEEC) over the 2005–2008 period. Contrary to previous studies, we estimated cost and revenue efficiency using Data Envelopment Analysis (DEA), which allowed us to focus on both the input and output sides of banks' efficiency. Our second stage analysis includes testing of the separability assumption and estimation of the truncated regression developed by Simar and Wilson (2007). We found evidence that: i) the size and financial capitalization of the bank are positively associated with cost and revenue efficiency; ii) foreign banks in CEEC are more cost efficient but less revenue efficient than domestic banks, suggesting different banking behavior between foreign and domestic banks; and iii) the loans-to-assets ratio was negatively associated with cost efficiency but positively associated with revenue efficiency, further stressing two different aspects of banking behavior in CEEC.

1. Introduction

It is now well established that the development of the financial, and particularly the banking system, promotes economic growth (Levin, 1997; Cetorelli and Gambera, 2001). It is not surprising that after the collapse of communism, the financial systems in Central and Eastern European Countries (CEEC) underwent drastic transformations. For example, in centrally planned economies, banks were state owned and strictly regulated, which prevented them from behaving efficiently. Financial entities were designed to perform activities that were in compliance with the central plan and in this manner support the apparatus of the central planner. After the collapse of communism, during the 1990s, restructuring (establishment of a two-tier banking system) and liberalization of banking systems started. Privatization of state-owned banks, entry of foreign banks, freeing of interest rates, changes in legislation (arrangements between debtors and creditors), and the establishment of prudential regulation and supervision were the most common tools of liberalization in CEEC (Nikiel and Opiela, 2002; Fries and Taci, 2005).

As capital markets in CEEC are still rather underdeveloped, the role of banking in the financial system appears to be even more substantial than it is in more developed countries. Therefore, the efficiency of banks is of particular interest both to regulators as well as to banks' management. For example, a regulator might be interested in knowing whether selling a domestic bank to foreign investors leads to

* The authors acknowledge the funding support of the Slovak Grant Agency for Science (VEGA Project No. 1/0393/12) and are indebted to comments from Tomáš Výrost and two anonymous referees.

higher efficiency, or more generally, what type of efficiency should be expected. The stability of the financial system might be of particular interest here. Are banks which bear more risk (in terms of the extent of loans and capital adequacy) more efficient? And if so, what type of efficiency is more typical of riskier banks? Similarly, bank owners and managers might be interested in the relationship between efficiency and profitability, as a specific type of efficiency is probably an outcome of a specific strategy.

Our study attempts to address several of these issues by analyzing a 2005–2008 sample of annual data of commercial banks in 11 CEEC: Bulgaria (BGR), Croatia (HRV), the Czech Republic (CZE), Estonia (EST), Hungary (HUN), Latvia (LVA), Lithuania (LTU), Poland (POL), Romania (ROU), Slovakia (SVK), and Slovenia (SVN). More specifically, we investigate the effect of bank size, the equity-to-assets ratio, profitability, ownership, and credit risk (loans-to-assets) on the cost and revenue efficiency of commercial banks in CEEC. Our study differs from previous studies in four important aspects. First, in contrast to most of the previous studies of bank efficiency in CEEC, we supplement a cost efficiency Data Envelopment Analysis (DEA) model with a revenue efficiency model. In most studies, cost efficiency is supplemented by profit efficiency, but as argued by Alcober et al. (2011, p. 2), “banks attempt not only to offer products and services at the minimum costs—i.e., to be cost efficient—but also to maximize revenues they generate—i.e., to be revenue efficient”. More interestingly, revenue maximization might be of greater importance for unsaturated markets, or countries with high economic growth rates, where the acquisition of market share might be more important than temporal cost inefficiency. Such situations imply behaviors other than cost saving, e.g., revenue maximization and revenue efficiency. Such perspectives are difficult to acquire using profit efficiency analysis. To our knowledge, only Iršová (2009) has attempted to measure revenue efficiency in CEEC. Second, we use a more homogeneous sample of commercial banks (excluding savings banks). Third, we use bias-corrected efficiency estimates calculated from an influential observation-adjusted sample. Fourth, in the second-stage analysis, we utilize the truncated regression developed by Simar and Wilson (2007), which accounts for the bias of the DEA efficiency estimates and their dependence. In contrast to Chortareas et al. (2011) and Chronopoulos et al. (2011), we test for the separability assumption and as a robustness check, we also perform our analysis under the multi-country, single-year frontier assumption.

2. Literature Review

Focusing on CEEC, Jemric and Vujcic (2002) used standard DEA models to measure bank efficiency in Croatia (1995–2000 data). They adopted the operating and intermediation approaches to bank analysis.¹ In the latter, fixed assets and software, the number of employees, and total deposits received were used as inputs, and

¹ The intermediation approach assumes that the bank acts as a financial intermediary between liability holders and debtors. Thus, a bank's role is to collect deposits and transform them using labor and capital in the form of loans (see Sealey and Lindley, 1977). This role is the dominant approach in the empirical literature. The production (operating) approach is an alternative that views banks as using labor and capital to produce deposits and loans. The choice of approach determines the types of inputs and outputs used in the empirical research.

total loans extended and short-term securities issued by official sectors were used as outputs. The authors found some evidence that the relationship between bank size and efficiency is U-shaped and that private and new banks are more efficient. Using a sample from 1997–2000, Nikiel and Opiela (2002) measured the efficiency of Polish banks using multi-year Stochastic Frontier Analysis (SFA). Their small sample allowed them to analyze the relationship between bank efficiency and customer type, where they found evidence that foreign banks with business customers are more efficient. In the first stage of their analysis, Hasan and Marton (2003) measured cost and profit (in)efficiency (using SFA and a multi-year frontier approach) in Hungary on a sample of data from 1993–1998. In the second stage of their analysis, the inefficiency estimates were regressed (using the OLS approach) on a set of bank-specific variables. Among other results, they found that in general, foreign banks were more cost and profit efficient, banks with higher equity ratios were less cost efficient, and larger banks were more efficient. This type of two-stage analysis has become standard in bank efficiency studies. Similar to Nikiel and Opiela (2002), Havrylchuk (2006) studied the efficiency of the Polish banking industry during the 1997–2001 period using DEA and an intermediation approach. Cost efficiency estimates were calculated for a separate (domestic and foreign banks) and common multi-year frontier. The inputs included capital, labor, and deposits, while the outputs included loans, government bonds, and off-balance sheet items. Surprisingly, neither the size of the bank nor capitalization was related to the efficiency estimates (in the second stage—a Tobit approach), but total loans to total assets were significant and negative, suggesting that banks that took more risks were less efficient.

Typically, single-country studies work with limited data (the case of CEEC). To mitigate data issues and allow for comparisons among countries, researchers have used multi-country data. With the exception of a few cases (see Poghosyan and Kumbhakar, 2010; Kořak and Zorić, 2011), one needs to work under the assumption of a common cross-country production frontier, i.e., it is assumed that banks in different countries can reach a common efficiency frontier² (a multi-country, single-year—MCSY—frontier). Typically, this frontier is also expanded across multiple years (a multi-country, multi-year—MCMY—frontier).

The literature is dominated by studies that employ SFA. Using a sample of annual data from 1997, Weill (2003) studied 47 banks in Poland and the Czech Republic. He found that foreign banks were more efficient. Fries and Taci (2005) studied the efficiency of banks across 15 East European countries. They used MCMY stochastic frontier estimation for the 1994–2001 period.³ Among other results, they found that private banks and banks that possess greater market power (market share of deposits) are more efficient than state-owned and other banks. Bonin et al. (2005) used similar data from 11 CEEC for the 1996–2001 period, but they performed a traditional two-stage analysis with SFA and OLS estimation. Foreign banks in 11 CEEC were more cost and profit efficient.⁴ Foreign and larger banks also proved to be more efficient in a study by Kasman and Yildirim (2006), who estimated

² Note that in an extreme case, such a frontier may be composed of bank(s) from a single country. To control for such cases, environmental variables are often used in the first or second stage of the analysis.

³ They also calculated single-year efficiencies, but because the final results were not sensitive to this choice, they reported results for the panel data only.

⁴ Evidence for profit efficiency was somewhat weaker.

an MCMY stochastic frontier for a sample of banks in eight CEEC (1995–2002 data). Mamatzakis et al. (2008) found that smaller banks were more profit efficient, while larger banks were more cost efficient. They argued that smaller banks operate in local markets and are therefore able to exercise some monopoly power, which in turn explains their higher profit efficiency. Again, a common MCMY stochastic frontier was estimated for the 1998–2003 period and 10 CEEC. Interestingly, some evidence of σ and β convergence in cost efficiency was also presented. Using a sample of 12 CEEC and data from the 1993–2000 period, Yildirim and Philappatos (2007) found that size, loans-to-assets, and equity-to-assets were all positively associated with cost efficiency (SFA with GLS fixed-effect estimation in the second stage).⁵

Grigorian and Manole (2006) used a standard DEA model to calculate an MCMY efficiency frontier for banks in 17 CEEC during the 1995–1998 period. In defining their sets of inputs and outputs, they relied on the value-added approach. Their findings from the second-stage analysis (censored Tobit) confirmed that foreign banks (strong evidence) and banks with higher equity-to-assets (weaker evidence) were more efficient. Stavárek (2006) included Portugal and Greece in an analysis of CEEC because he considered both countries to be among the least efficient in the then-current European Union (EU) countries, while the rest of the CEEC were aspiring EU members (the countries were grouped as follows: i) POR, GRE, ii) CZE, HUN, POL, SVK, iii) EST, LVA, LTU, iv) BGR and ROU). His analysis included three years, 2001–2003, and to calculate the efficiency scores, he adopted the intermediation approach and a standard DEA with an MCMY frontier. Three inputs (total personnel costs, total deposits, and fixed assets) and two outputs (total loans and net interest income) were considered. As expected, banks in the first group were the most efficient, followed by the second, third, and fourth groups. In a subsequent analysis, he also found that larger banks were more efficient, but his evidence suggests that the differences have decreased. For a sample of 22 EU countries (including several CEEC) during the 2000–2008 period, Chortareas et al. (2011) found that larger, foreign-owned banks and banks with higher equity-to-assets ratios were more efficient.⁶ However, these results are more difficult to compare with previous studies, as the CEEC were combined in a single MCMY DEA model with far more developed EU countries. Under the intermediation approach, personal expenses, total fixed assets, deposits, and short-term funding were used as inputs, and total loans, total other earning assets, and free-based income were used as outputs. Chronopoulos et al. (2011) used a set of CEEC and a period (2001–2007) that are similar to our study. They used the intermediation approach and MCMY cost and profit efficiency DEA models. In the second stage of their analysis, they found that: i) larger banks tended to be more cost and profit efficient, ii) banks with higher equity-to-assets ratios were less cost efficient, iii) more diversified banks were more cost and profit efficient, and iv) foreign ownership did not increase bank efficiency. Both Chortareas et al. (2011) and Chronopoulos et al. (2011) used truncated regressions with bootstrapped standard errors (in their second-stage analysis), as was

⁵ They also used the Distribution Free Approach, but because it is used very rarely, we do not review this approach.

⁶ The same conclusions may also be drawn for a sub-sample of 16 countries; see models 7–8 in Table 4 of Chortareas et al. (2011).

recommended in Simar and Wilson (2007). Finally, Iršová (2009) used both the SFA and DEA approaches while calculating cost and revenue efficiencies for five CEEC during the 1995–2006 period. In the second stage of her analysis, she found some evidence (using both Tobit and OLS) that the equity-to-assets ratio is positively related to revenue efficiency.⁷

Two in-depth surveys on the measurement of bank performance have been published recently, namely, Fethi and Pasiouras (2010) and Liu et al. (2012). Fethi and Pasiouras (2010) provided a survey of 196 studies on bank performance. They noted that DEA is the most widely applied operational research technique. As stated previously, the first stage of calculating DEA efficiency scores (as performance measures) is often supplemented by a second stage in which the determinants of efficiency are evaluated. The hypothesized determinants and estimation techniques vary across studies. The typical hypothesized explanatory variables include event indicators (e.g., EU accession and the starting dates of banking reforms) and bank- and country-specific factors. Liu et al. (2012) surveyed the DEA literature and identified two-stage DEA analysis as one of the five most active DEA sub-areas in recent years, with two influential studies: Simar and Wilson (2007) and Banker and Natarajan (2008). In earlier studies, the second-stage analysis had been performed using Tobit, OLS, GMM, and GLS regressions. McDonald (2009) has argued that efficiency scores are not the outcome of a censoring process but are instead fractional data. In this context, for the conditional mean of the fractional response that restricts the predicted values to within the unit interval, some studies have employed Papke and Wooldridge's (1996) approach, which was advocated by both Banker and Natarajan (2008) and McDonald (2009).⁸ Simar and Wilson (2007, 2011) have argued that the efficiency variable follows a truncated distribution and that a truncated regression is therefore a good methodology to ensure efficient estimation of the second-stage estimators. Perhaps more importantly, the previous methods have not accounted for the obvious correlations among efficiency scores, while the bootstrap techniques of Simar and Wilson (2007) take the dependence structure of efficiency scores into account. Therefore, we decided to use their approach. However, in applied research, the diagnostic checks required for the proper estimation of the Simar and Wilson (2007) regression are not performed frequently.⁹

The rest of the paper is organized as follows. In the next section, we specify the data, how it was cleaned, and the adjustments made to the data and samples used in this paper. Section 4 describes the methodologies employed to: i) calculate efficiency scores and ii) perform the second-stage regression. In Section 5, we present the results on cost and revenue efficiency and the second-stage regression results. Finally, Section 6 summarizes our findings.

3. Data

The geographical coverage of this study is as follows: Bulgaria (BGR), Croatia (HRV), Czech Republic (CZE), Estonia (EST), Hungary (HUN), Latvia (LVA),

⁷ However, in the DEA approach employed, she calculated MCSY frontiers, which were stacked into a panel in the second-stage analysis.

⁸ This approach has also been used in Chronopoulos et al. (2011) and Chortareas et al. (2011) as a sensitivity "check" of the results.

⁹ Even outside bank efficiency studies.

Lithuania (LTU), Poland (POL), Romania (ROU), Slovakia (SVK), and Slovenia (SVN). Our sample of annual data consists of 665 observations for 187 different commercial banks over the 2005–2008 period. These data are measured as the averages of each variable between January 1 and December 31 to avoid large changes during the year. The data were acquired from financial statements as provided in the Bankscope database by Bureau van Dijk. We relied largely on Bankscope's definitions of the variables, homogenized into a global format.

The Bankscope database was supplemented by other data sources, e.g., a list of financial institutions provided by the European Central Bank, lists of domestic banks and branches of credit institutions prepared by the relevant central banks, and annual reports provided by individual banks helped us to compile more precise data.¹⁰ First, the initial sample consisted of all banks operating in the examined countries. Central banks, investment banks, savings banks, cooperative banks, real estate and mortgage banks, specialized state financial institutions, and branches of banks were subsequently excluded from the sample. Therefore, we constructed a list of commercial banks in each country. It should be noted that not all of the commercial banks operated over the entire time period. Second, to be included in the final sample, the bank had to have all the variables available for the given year. We opted to employ unconsolidated data whenever possible. When these data were not available, we used consolidated data instead. In most cases, the data were drawn from financial reports prepared under International Financial Reporting Standards (IFRS). In other cases, data from financial statements correspond to those prepared under Generally Accepted Accounting Principles (GAAP).¹¹ All data are reported in EUR, converted to the value of the EUR in 2008 and adjusted for inflation using Gross Domestic Product (GDP) deflators.¹² These adjustments were performed to increase data comparability.

Table 1 summarizes the number of banks available for this study. Note that Chronopoulos et al. (2011) also included savings banks in their sample, which should lead to a larger number of banks. However, this was not the case for Hungary (see Table 1 in Chronopoulos et al., 2011). We are unable to explain this data discrepancy. Thus, note that it is difficult to compare the datasets used in previous studies with our somewhat restricted dataset. However, we believe that the cleaning and adjustment of the data helped us form a more homogeneous sample of banks. In an interesting paper, Iršová and Havránek (2010) observed that in efficiency studies, commercial banks tend to have significantly different (typically lower) efficiency scores than other types of banks. Therefore, they recommend that samples should be more homogeneous with respect to bank type.

Many studies on bank efficiency have been conducted since the seminal work of Greenbaum (1967), but there is still no consensus on which inputs and outputs to use. Berger and Humphrey (1997) summarized and critically reviewed different

¹⁰ The Bankscope classification for commercial banks is not necessarily the same as the definitions used by central banks. For example, commercial and savings banks can be pooled under the common label "commercial banks". Researchers need to pay attention to this issue.

¹¹ Following Maudos and de Guevara (2004) and Lozano-Vivas and Pasiouras (2010), among others.

¹² The data on GDP deflators (with 2000 being the base year) were acquired from International Monetary Fund (IMF) data sources.

Table 1 Number of Banks

	BGR	CZE	EST	HRV	HUN	LVA	LTU	POL	ROU	SVK	SVN	Total
2005	17	11	5	24	15	17	6	24	21	10	16	166
2006	18	11	6	25	16	16	6	24	22	11	14	169
2007	19	12	6	24	15	15	6	24	22	11	15	169
2008	19	12	6	25	12	14	6	22	21	10	14	161
2005–2008	73	46	23	98	58	62	24	94	86	42	59	665

approaches for measuring efficiency employed in 130 studies. We adopted the intermediation approach and assumed that banks produce two outputs: total loans and other earning assets. The prices of those outputs are represented by the ratios of interest received on loans to total performing loans and noninterest income to other earning assets, respectively. Total deposits and total costs represent the two inputs. The prices of those inputs are total interest expenses to total deposits and total costs to total assets, respectively.¹³

To examine the determinants of bank efficiency, we selected the following explanatory variables, which have been used in most studies on bank efficiency:

- (a) The average capital ratio (ETA) is measured as the bank's equity over total assets. A lower ETA should lead to lower efficiency scores, as less equity implies higher risks being taken at greater leverage. More capitalized banks should be able to absorb losses on loans much more easily than less capitalized banks.
- (b) The profitability ratio is defined as the return on average equity (ROAE). One can expect more efficient banks to earn higher profits, which should lead to a positive relationship between ROAE and efficiency.
- (c) Additionally, to account for the size of each bank and its possible effects on efficiency, we used the natural logarithm of total assets (LNTA).
- (d) The loan risk ratio (LOANS) is measured as total loans to total assets. Loans are generally a major item on a bank's balance sheet. LOANS is expected to be positively related to the risk of bank failure. This ratio can be accompanied by higher costs and lead to lower efficiency.
- (e) Foreign ownership (FOREIGN) corresponds to a dummy variable that takes a value of 1 if more than 50% of a bank's total stock of shares is held by foreign shareholders. This information was obtained from the 2008 edition of the Bankscope database. Under the assumption that foreign investors will transfer their knowledge, skills, and technology, these banks are expected to be more efficient and perform better than those with less than 50% foreign participation

¹³ Following Casu and Molyneux (2003), Hahn (2007), and others, we specified an intermediation-oriented model. One of the reviewers pointed out that the prices of inputs are not independent and therefore another model specification should be considered, for example, next to the price of total deposits, the price of operating capital, equal to the ratio of operating expenses to fixed assets (see Berger et al., 2009). We are thankful for this comment. In our specification, the correlations between input prices as well as the products of inputs and their prices were low (0.0229 and 0.1152, respectively). Further on, we decided to use an overall measure of costs, as such a measure makes comparison among different types of commercial banks more meaningful.

Table 2 Descriptive Statistics of the Entire Dataset

	2005, n = 166		2006, n = 169		2007, n = 169		2008, n = 161		Multi-year, n = 665	
	Mean	St. Dev.	Mean	St. Dev.						
Total Loans (TL)	1,271.288	1,998.446	1,581.913	2,409.169	1,993.279	2,959.488	2,352.064	3,424.485	1,795.374	2,775.275
Other Earning Assets (OEA)	962.542	1,992.315	1,067.120	2,147.630	1,110.262	2,199.790	1,129.738	2,284.148	1,067.139	2,158.718
Total Deposits (TD)	1,965.398	3,281.628	2,312.048	3,790.666	2,699.905	4,274.174	3,010.609	4,675.033	2,493.209	4,053.099
Total Costs (TC)	168.037	297.591	185.279	363.159	215.539	401.321	261.609	457.437	207.145	385.297
Price of TL	0.124	0.069	0.112	0.056	0.104	0.051	0.106	0.042	0.112	0.056
Price of OEA	0.030	0.054	0.037	0.073	0.041	0.105	0.051	0.151	0.040	0.102
Price of TD	0.044	0.076	0.068	0.370	0.079	0.511	0.063	0.217	0.063	0.338
Price of TC	0.077	0.046	0.072	0.060	0.071	0.061	0.076	0.064	0.074	0.058
LNTA	13.627	1.579	13.814	1.607	14.064	1.533	14.196	1.523	13.923	1.576
LOANS	53.792	17.373	56.900	16.988	60.017	17.079	62.818	16.454	58.349	17.309
ETA	11.980	6.755	11.623	6.666	10.878	5.763	10.747	5.435	11.311	6.208
ROAE	13.748	12.043	14.305	15.331	14.109	15.435	6.811	27.254	12.302	18.601

Notes: TL, OEA, TD, and TC are reported in thousands of EUR. The prices of TL, OEA, TD, and TC are per unit. LNTA is the natural logarithm of total assets. LOANS, ETA, and ROAE are reported in percentages. Turning to the variables used in the estimation of the DEA frontiers (i.e., inputs, outputs, and the prices of inputs and outputs), the initial sample included six negative values of the variable "price of other earning assets". We decided to replace these values with 0.0001. This influences the revenue efficiency estimates only. At the recommendation of one of the reviewers, we re-calculated revenue efficiencies replacing the prices of OEA with the following variable: $\text{Price of OEA} + \lfloor \min(\text{Price of OEA}) \rfloor + \varepsilon = \text{Price of OEA} + 0.02285$ (small $\varepsilon > 0$). This ensured positive values. The correlation of the former and the resulting revenue efficiency scores was 0.9547. We therefore decided to work with the former results. Regarding the explanatory variables (i.e., LNTA, LOANS, ETA, ROAE), it should be noted that we recorded 45 negative values of the variable ROAE. Where this variable was used in the separability assumption tests, the negative values were replaced with 0.0001, as these variables are included in the DEA model.

(domestic banks). Moreover, it is believed that foreign banks in transitional countries provide greater financial stability (through higher liquidity in times of crises, e.g., Dinger, 2009) and that their presence also influences the behavior of other banks in the country (e.g., Lehner and Schnitzer, 2008).

Finally, country dummy variables were used to control for the country-specific effects of the banks in the sample. Where possible, the second-stage analysis also included year dummies and country-specific variables, which should mitigate differences in production technologies: GDP per capita based on purchasing power parity (GDP/C), real GDP growth (Δ GDP), the GDP deflator (INFLATION), the index of financial freedom (IFF), and the Hirschman-Herfindahl index calculated from total deposits (HHI). A more detailed explanation of the variables is provided in *Appendix A*. Descriptive statistics for inputs, outputs, their prices, and bank-level determinants (except for the dummy variable FOREIGN) considered in this study are presented in *Table 2*. Descriptive statistics for macroeconomic and regulatory-level determinants are not shown in order to conserve space.

4. Methodological Issues

The results of previous studies appear to be sensitive to the choice of the models used in the first and second stages of the analysis. Thus, we describe our choices and procedures in greater detail. There are two main objections to the non-parametric DEA method. First, it is argued that the DEA frontier is deterministic and unable to account for measurement errors (e.g., Weill, 2003; Fries and Taci, 2005; Yildirim and Philappatos, 2007). Second, this approach is sensitive to outliers (e.g., Havrylychuk, 2006). However, by calculating the bias-corrected efficiency estimates from a sample adjusted for influential observations, both objections can be addressed. Therefore, we consider the non-parametric DEA method suitable and decided to use it in our study.

4.1 The DEA Frontier

According to Cooper et al. (2007), DEA is a linear programming methodology for measuring the relative efficiency of peer production units, which are termed “Decision-Making Units” (DMUs) and have multiple outputs and inputs.¹⁴ We applied DEA techniques to identify the cost and revenue efficiency scores for each bank within an MCSY and MCMY framework. The traditional cost efficiency model can be traced back to Farrell (1957) and Debreu (1951), who assumed that the unit costs of inputs are identical among DMUs. To be cost efficient, the DMU must be both technically efficient (adopting the best practice technology) and allocatively efficient (selecting the optimal mix of inputs to minimize the costs for a given output). Revenue efficiency is analogous to cost efficiency (the product of technical and allocative efficiency); however, one strives for revenue maximization for a given set of inputs. Due to imperfect competition in actual markets and the lack of common input and output prices among DMUs, we decided to follow Tone (2002) when calculating cost and revenue efficiency.

We define \mathbf{y}_o as the $s \times 1$ vector of the o -th bank’s s outputs, \mathbf{x}_o is the $m \times 1$ vector of its m inputs ($k = 1, \dots, m$), and \mathbf{Y} is the $s \times n$ matrix of outputs (n denotes the number of DMUs). Next, consider matrix $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_n)$, the columns of which

¹⁴ The term “Data Envelopment Analysis” was first used in a paper by Charnes et al. (1978).

are vectors $\bar{x}_i = (c_{1i}x_{1i}, \dots, c_{mi}x_{mi})^T, \forall i = 1, \dots, n$, where x_{ki} denotes the k -th input of the i -th DMU and c_{ki} is the corresponding price of the k -th input of the i -th DMU. Additionally, e is a $1 \times n$ row vector with all elements being equal to 1, and λ denotes an $n \times 1$ vector of weights. Following Tone (2002), the new cost efficiency linear programming (LP) problem can be defined as:

$$\bar{e}\bar{x}_0^* = \min_{\bar{x}, \lambda} \bar{e}\bar{x} \quad (1)$$

subject to $\bar{x} \geq \bar{X}\lambda$
 $y_0 \leq Y\lambda$
 $e\lambda = 1$
 $\lambda \geq 0$

The new cost efficiency of a DMU_o is calculated as:

$$NCE_o = \bar{e}\bar{x}_0^* / \bar{e}\bar{x}_0 \quad (2)$$

where \bar{x}_0^* is the optimal solution of the LP given above and $\bar{e}\bar{x}_0$ is the observed value for DMU_o . The cost efficiency score is a ratio bounded at zero and one, where one indicates that the bank is cost efficient (a technically and allocatively efficient unit). The revenue efficiency model is directly analogous to the cost efficiency case, and therefore, a detailed description is not provided here.¹⁵ The constraint $e\lambda = 1$ specifies that we assumed Variable Returns to Scale (VRS) for the frontier, which is more common in bank efficiency studies compared to the Constant Returns to Scale (CRS) assumption. Even if this assumption is violated and the true frontier is governed by CRS, the consequences are not dramatic (loss of efficiency but not consistency of the efficiency scores; see Simar and Wilson, 2002).

Clearly, DEA efficiency estimates are sensitive to the presence of influential observations on the efficiency frontier.¹⁶ These influential observations may be the result of (i) measurement errors, (ii) DMUs with exceptional performance, or even (iii) unknown heterogeneity in the DMUs (an observation from a different data generating process; Simar, 2003). We consider the latter two to be relevant to our data. Unfortunately, it is difficult to distinguish between them (see Yang et al., 2010 for a brief discussion).

In the literature on DEA, there are several approaches for detecting influential observations.¹⁷ For example, the approach of Wilson (1993, 2010) is performed on inputs and outputs but not on efficiency scores¹⁸ (for applications, see, e.g., Jamasb et al.,

¹⁵ A cost efficiency model is an input-oriented model, as it minimizes inputs at a given level of outputs, i.e., assessing a bank's ability to control costs. A revenue efficiency model is an output-oriented model that maximizes revenues for a given set of outputs, i.e., assessing a bank's ability to generate revenues. For a brief survey of DEA models, see Coelli et al. (2005).

¹⁶ Within the DEA framework, we use the term "influential observation" instead of "outlier", as it provides a better description of the interrelatedness of efficiency estimates.

¹⁷ A different approach not considered here would be to adjust DEA methods to be robust against outliers in a similar vein as in Cazals et al. (2002).

¹⁸ Wilson's (1993) approach is based on the previous work of Andrews and Pregibon (1978) and was considered rather computationally intensive. As noted in Wilson's (2010) comment, this drawback is mitigated with increasing computational power.

2008; Benito-Lopéz et al., 2011). Wilson's (1993) approach is a method for identifying multivariate outliers which can be used without any reference to DEA. In contrast to frontier-dependent approaches, a priori assumptions regarding the correct model specification are not necessary. With frontier-dependent models, it is not surprising that with different model specifications (which define the frontier), one typically finds different influential observations. As we considered only DEA models with VRS frontiers, this issue is not relevant to our study. We therefore decided to obtain a more homogeneous dataset by removing influential DMUs using the frontier-dependent method of Sousa and Stošić (2005). After the influential observations (DMUs) were identified, they were removed from the samples, and the corresponding DEA models were re-estimated using these influential observation-adjusted samples. A detailed description of the Sousa and Stošić (2005) procedure is provided in *Appendix B*.

All DEA estimators are biased by construction (Simar and Wilson, 2000). The appropriate approach to correct for this bias depends on whether the independence assumption holds. If independence does not hold, for an input (output)-oriented model, there are production spaces with a specific mix of inputs (outputs) and corresponding levels of outputs (inputs) at which DMUs are more efficient (Simar and Wilson, 1998; Wilson, 2003). For example, banks might be efficient only within some range(s) of input (output) values. Technically, this independence assumption may be perceived as too restrictive (see Wilson, 2003, p. 362, for a discussion). By rejecting this assumption, we should use the heterogeneous bootstrap developed by Simar and Wilson (2000). Otherwise, if the independence assumption is economically meaningful, one can use the homogeneous bootstrap presented in Simar and Wilson (1998). To test the independence assumption, we followed Wilson (2003).¹⁹ Finally, bias correction of the efficiency estimates was performed using the heterogeneous bootstrap presented in Simar and Wilson (2000), and the resulting bias-corrected estimates $\hat{\theta}_i$ (from the samples adjusted for influential observations) were used in the second-stage regression.²⁰

4.2 Second-Stage Estimation

Regardless of whether one adopts the model considered by Simar and Wilson (2007) or another model, “*one should carefully consider what restrictions are necessary, and whether these are reasonable. Ideally, restrictions should be tested*” (Simar and Wilson, 2011, p. 216). Assumptions A1–A8 in Simar and Wilson (2007 p. 34–37) define a semi-parametric data-generating process (DGP) that generates observations (inputs, outputs, and environmental variables used in the regression analysis). Assumptions 1 and 2 specify two related issues.

In this instance, the assumption of non-independence (A1) means that the vector of environmental variables \mathbf{z} is not independent with respect to (\mathbf{x}, \mathbf{y}) , where \mathbf{x} and \mathbf{y} denote vectors of inputs and outputs, respectively. This assumption may be tested by the methods described in *Appendix C*. However, these tests were

¹⁹ See *Appendix C* for statistics and the calculation of critical values.

²⁰ The bias correction was performed only on efficiency scores with sample variance satisfying Eq. (50) in Simar and Wilson (2000). The algorithm for the heterogeneous bootstrap can be found in Simar and Wilson (2000).

not performed because if the second-stage analysis finds significant relationships, it seems reasonable to assume that this assumption of non-independence between the input-output space and the environmental variables is not violated.

The separability assumption (A1 and A2) states that the exogenous variables used in the second stage of the DEA should not affect the efficiency frontier but are only allowed to affect the efficiency scores. If this assumption is violated, it seems reasonable to include those variables in the set of inputs or outputs. Otherwise, the estimated coefficients in the second-stage regression will be biased. Simar and Wilson (2007) recommended testing this assumption using the statistics defined in Simar and Wilson (2001), which can be used to test for irrelevant inputs or outputs. We adjusted these tests slightly to use them in our cost and revenue efficiency models. Note that we considered ETA and LNTA as possible inputs and LOANS and ROAE as outputs. The procedures are described in *Appendix D*.

Finally, we estimated the truncated regression for evaluating the dependence between the efficiency scores and explanatory variables, following the approach of Simar and Wilson (2007). Specifically, in this study, the following truncated regression was estimated:

$$\hat{\theta}_i = \alpha + \mathbf{Z}_i\boldsymbol{\beta} + \varepsilon_i \quad (3)$$

where α is a constant term and \mathbf{Z}_i is a vector of explanatory variables for bank i , including bank- and country-specific variables, which is hypothesized to affect the bias-corrected efficiency score ($\hat{\theta}_i$) of bank i . $\boldsymbol{\beta}$ is a vector of parameters to be estimated. ε_i is an error term assumed to be $N(0, \sigma_\varepsilon^2)$ distributed with left truncation at $(-\mathbf{Z}_i\boldsymbol{\beta})$ and right truncation at $(1-\mathbf{Z}_i\boldsymbol{\beta})$. The set of explanatory variables includes: i) the bank-specific variables of interest to this study: LNTA, ETA, LOANS, ROAE, and FOREIGN, ii) country-specific variables: country dummy variables, time dummy variables, GDP/C, Δ GDP, INFLATION, IFF, and HHI.

Simar and Wilson (2007) proposed two bootstrap algorithms that can be used to derive the standard errors of the parameter estimates. As our dependent variables are already bias-corrected efficiency scores, we used Algorithm 1 (see *Appendix E* for details).

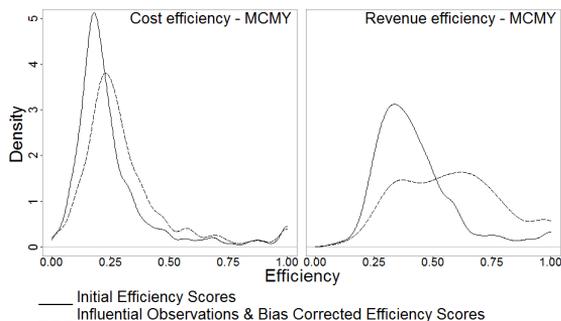
5. Empirical Results

5.1 Efficiency Levels

Before calculating the bias-corrected efficiencies, we removed influential observations, as suggested by the tests developed by Sousa and Stošić (2005). For both MCMY samples, only the removal of one DMU was required (for the MCSY samples considered in Section 5.3, up to two DMUs were removed). Next, the independence tests suggested that the correct approach for bias correction was to use the heterogeneous bootstrap.²¹ These results have an interesting implication, as they

²¹ The test statistics (C.2) and (C.3) were 0.088 and 0.002 for the CE and 0.099 and 0.002 for the RE models, respectively, and were all significant at the $< 1\%$ significance level. For the samples considered in Section 5.3, the data for the 2006 and 2007 samples and the RE model were not significant at the $< 10\%$ significance level. However, even in those cases, the heterogeneous bootstrap may be applied.

Figure 1 Kernel Density Estimates (Gaussian Smoothing Kernels) of Efficiency Scores



suggest that the conditional distribution of efficiency scores (conditional on the inputs and outputs considered in our DEA models) is not the same as the unconditional distribution. Therefore, there may be specific input (output) spaces for which higher efficiency is more probable. In *Figure 1*, we plotted the distribution of the efficiency scores calculated from a sample without influential observations and with bias correction (dashed line). The differences in the distributions can primarily be attributed to the removal of influential observations, as few of the efficiencies required bias correction (see Footnote 20). From *Figure 1*, it also seems clear that influential observations can have significant impacts on other efficiency scores because they naturally lie on the efficiency frontier. If these observations are removed, the efficiency frontier moves, and as a result, non-efficient banks may move toward the efficiency frontier, i.e., their efficiency scores may change. It is also interesting to observe that while for the cost efficiency model, the removal of an influential observation caused the distribution of efficiency to be shifted to the right, in the revenue efficiency model, the shape of the distribution also changed. This observation further emphasizes our argument for taking influential DMUs into account. In what follows, we will work only with influential observations and bias-corrected efficiency scores.

We observed no dramatic changes in the average cost and revenue efficiencies during the 2005–2008 period, although cost efficiency declined slightly and revenue efficiency increased (see Panel A in *Table 3*). As in previous studies, notable differences among countries were observed (see Panel B in *Table 3*).²²

The average efficiencies are low when compared to other studies. We re-estimated the DEA models with a different specification input specification, using total deposits and other costs (noninterest expenses, personnel expenses) instead of total deposits and total costs. The means were still very low—0.267 for CE and 0.227 for RE (MCMY frontier). The removal of influential observations was based on a threshold provided by Sousa and Stosic (2005). But as we tried several specifications of

²² A straightforward application of the *t*-test is not recommended, owing to the dependence of the efficiency scores. Our assessment of the differences is certainly subjective. Note that together with the reported standard deviations and our knowledge of the efficiency score's boundaries of zero and one, we can at least form a sensible opinion about the extent of the differences.

Table 3 Cost and Revenue Efficiency Estimates of the MCMY Frontier

Panel A	Cost efficiency				Revenue efficiency			
	2005	2006	2007	2008	2005	2006	2007	2008
Mean	0.336	0.339	0.304	0.263	0.421	0.457	0.463	0.495
St. Dev.	0.222	0.206	0.184	0.168	0.179	0.200	0.194	0.208
Median	0.272	0.280	0.257	0.220	0.381	0.400	0.410	0.440
Panel B	Mean	St. Dev.	Median		Mean	St. Dev.	Median	
SVK	0.366	0.180	0.326		0.496	0.146	0.463	
CZE	0.445	0.368	0.352		0.470	0.279	0.410	
POL	0.332	0.182	0.279		0.535	0.198	0.517	
HUN	0.196	0.089	0.187		0.561	0.224	0.577	
SVN	0.286	0.073	0.282		0.430	0.139	0.396	
BGR	0.308	0.144	0.261		0.494	0.210	0.416	
ROU	0.167	0.066	0.171		0.441	0.166	0.408	
LTU	0.392	0.154	0.379		0.381	0.162	0.320	
EST	0.572	0.312	0.455		0.529	0.236	0.442	
LVA	0.429	0.226	0.376		0.363	0.142	0.314	
HRV	0.260	0.087	0.242		0.372	0.150	0.315	
Total	0.311	0.198	0.258		0.459	0.197	0.410	

their approach (see the end of *Appendix B*), we do not think this is the issue. Further on, it should be noted that even if we remove influential observations, it does not necessarily increase the average efficiency. We believe that the low efficiencies are due to the input specification, most notably because we use only two inputs and two outputs. Usually, higher dimensions lead to more efficient units, which drive the averages upwards.

The average cost efficiency was higher for the Baltic countries and the Czech Republic. Lower values were observed for Romania and Hungary. However, the differences between the average revenue efficiencies were smaller, with Hungary having the largest score, while two Baltic States (Lithuania and Latvia) were further from the efficiency frontier. This observation may suggest different banking behaviors for specific countries and banks (to partially control for this possible heterogeneity, our second-stage models include country-specific variables). Thus, we calculated rank-based correlations between the cost and revenue efficiencies for each year at the bank level. The correlations were all positive and, except for 2005, were all significant. Similar results can be found in Iršová (2009).²³ However, the correlations found here were much lower than those between the cost and profit efficiencies reported in Chronopoulos et al. (2011). These lower correlation coefficients can be viewed as an indicator of different (although not unrelated) banking behaviors.

²³ $\rho_{2005} = 0.111$, $\rho_{2006} = 0.277^{***}$, $\rho_{2007} = 0.310^{***}$, $\rho_{2008} = 0.333^{***}$ for the Spearman rank correlation and $\rho_{2005} = 0.072$, $\rho_{2006} = 0.186^{***}$, $\rho_{2007} = 0.211^{***}$, $\rho_{2008} = 0.233$ for the Kendall tau-b correlations, where *** denotes the 1% significance level. In Iršová (2009), the Spearman rank correlations between cost and revenue DEA models ranged from 0.05 to 0.46.

Table 4 Truncated Regression of Simar and Wilson's (2007) Analysis of Cost and Revenue Efficiencies

	Cost Efficiency				Revenue Efficiency			
	Model 1		Model 2		Model 1		Model 2	
	coef	std	coef	std	coef	std	coef	std
intercept	-0.1528	0.0627	0.1085	0.2148	-1.1685 **	0.0532	-1.0723 ***	0.1910
LNTA	0.0300 **	0.0042	0.0342 **	0.0040	0.1001 **	0.0036	0.0990 ***	0.0036
ETA	1.0256 **	0.0986	1.0454 **	0.0957	1.2002 **	0.0848	1.2004 ***	0.0844
LOANS	-0.2263 **	0.0302	-0.2018 **	0.0300	0.1432 **	0.0268	0.1398 ***	0.0266
ROAE	0.0005	0.0002	0.0001	0.0002	0.0012 **	0.0002	0.0014 ***	0.0002
FOREIGN	0.0310 *	0.0112	0.0286 *	0.0108	-0.0369 **	0.0099	-0.0364 ***	0.0099
Year Dummies	NO		YES		NO		YES	
Country Dummies	YES		YES		YES		YES	
GDP/C			-0.0000 *	0.0000			0.0000	0.0000
Inflation			-0.8228 *	0.3508			0.2524	0.3025
IFF			0.0019	0.0012			0.0004	0.0011
HHI			-0.3289	0.3807			0.0036	0.3788
Δ GDP			0.9957 **	0.2436			-0.4173 *	0.2153
No. of observations	664		664		664		664	
pseudo R^2	40.52%		44.08%		58.21%		58.88%	

Notes: The table presents estimates based on Simar and Wilson's (2007) bootstrap procedure. The heading coef stands for estimates of the regression coefficients and std for the standard errors of these coefficients. The results are rounded to the fourth decimal place. For the cost efficiency model, the coefficient of GDP/C is -0.0000217 (0.0000099 std. error), and for the revenue efficiency model, the results are -0.0000053 (0.0000088). The dependent variable is the cost and revenue efficiency score. The pseudo R^2 was calculated as the square of the Pearson's correlation between the predicted and actual values of the efficiency scores. *, **, *** indicate variables significant at the 10%, 5%, and 1% levels, respectively.

5.2 The Relationship Between Efficiency and Bank-Specific Variables

To explain the variability in the cost and revenue efficiencies, we regressed the efficiencies on the set of relevant bank- and country-specific variables. *Table 4* reports the regression results for four models. However, the separability assumption was tested first. At the $< 5\%$ significance level, we were unable to reject the null hypothesis. Thus, our data suggest that LNTA, ETA, ROAE, and LOANS do not change the production frontier significantly (although they are allowed to influence the efficiency scores).²⁴ For both cost and revenue efficiency we considered two models. Model 1 considers only country-specific dummy variables, while Model 2 also includes other country-specific environmental variables. The significance and signs of the bank-specific variables proved to be insensitive to the inclusion of country-specific variables in the regression.

²⁴ The D.3 and D.4 test statistics (specified in *Appendix D*), including the p -values in parentheses, were 128.467 (0.986) and 401.115 (0.894) for the cost efficiency model and 126.552 (0.236) and 109.424 (0.090) for the revenue efficiency model.

5.2.1 Determinants of Cost Efficiency

The coefficient on the size (LNTA) of the banks was positive and significant. This finding is not consistent with Havrylchuk (2006) but is consistent with other studies covering CEEC (e.g., Stavárek, 2006; Chortareas et al., 2011; Chronopoulos et al., 2011). This result was expected, as due to their size, larger banks can attain lower unit costs. Furthermore, there might also be an effect due to increasing returns to scale (see Hauner, 2005). The equity-to-assets (ETA) ratio was found to be positive and significant, similar to the findings of Grigorian and Manole (2006) and Chortareas et al. (2011) but not to those of Chronopoulos et al. (2011). According to the literature, there are two potential reasons for this relationship. First, higher equity alleviates agency problems between the management and owners. As owners acquire higher stakes in the bank, their tendency to monitor the management is higher, which leads to higher cost discipline and thus higher cost efficiency (see Mester, 1996, and Eisenbeis et al., 1999, for further details). The loans-to-assets ratio (LOANS) was significant and negative, suggesting that on average, banks with higher loans-to-assets ratios were less cost efficient. This finding might be the result of holding riskier loans or having poor credit management. The proxy for a bank's profitability (ROAE) was positive but not significant. The last variable of interest was the variable capturing the ownership (domestic/foreign) of the banks (FOREIGN), which was positive and significant, similar to the results of Grigorian and Manole (2006) and Chortareas et al. (2011) but not to those of Chronopoulos et al. (2011). This variable is of special interest mostly for transitional countries, where liberalization efforts consist partly in privatizing state-owned banks (often to foreign banks). Our results suggest that foreign banks are more cost efficient in CEEC.

Real GDP growth was positive and significant, which was not surprising. In general, as the economy grows and the output of the bank increases, per-unit fixed costs tend to decline, thus increasing cost efficiency. This finding is similar to that obtained by Yildirim and Philappatos (2007). Interestingly, the coefficient on GDP per capita was negative (and significant), indicating that banks operating in more developed economies were less cost efficient. The coefficient on inflation was significant and negative, which supports the general view that inflation hinders creditor institutions. The coefficient on HHI was negative, which was expected, as one can assume that with higher concentration, there is less bank competition, which in turn leads to less efficient banking. Such markets are believed to be detrimental to debtors, particularly in transitional economies, where non-bank sources of financing are limited (e.g., Wieneke and Gries, 2011). However, the significance of this relationship has not been proven. Note that the inclusion of these variables did not change the coefficients on bank-specific variables and that it increased the *pseudo R*² only marginally.

5.2.2 Determinants of Revenue Efficiency

The coefficients on size and the equity-to-assets ratio were positive and significant, similar to the results of the cost efficiency regressions. Hauner (2005) argued that compared to smaller banks, larger banks might have larger output-to-input ratios—an effect of increasing returns to scale. A positive effect of equity-to-total assets was also found in Iršová (2009). For the revenue efficiency model, this result might be explained by the argument that banks with higher efficiency “will

have higher profits and hence will be able to retain more earnings as capital” (Carvallo and Kasman, 2005, p. 70). However, it also seems plausible to assume that banks with higher equity-to-assets ratios have higher protection and can engage in riskier, more rewarding investment projects. There were two main differences between the cost and revenue regressions. First, the FOREIGN variable was significant (as before) but negative, indicating that foreign banks are less revenue efficient. Similarly, the loans-to-assets ratio was significant but positive (contrary to the negative sign in the cost efficiency regression). Banks that hold relatively more loans have higher revenue efficiency. These results also suggest that banks in CEEC that had higher loans-to-assets ratios were less cost efficient (possibly due to lower standards in credit risk management due to the different strategy employed by these banks), but they were also more revenue efficient, as they probably gained a greater market share relative to their asset sizes. Profitability was not significant for cost but was significant for revenue efficiency scores, and the coefficient was positive as expected. This finding naturally suggests that at least in the short term, the importance of revenue maximization might be higher (in terms of the commercial success of banks) than that of cost minimization. Interestingly, none of the relevant environmental variables were significant at the < 5% significance level.

The effect of loans-to-assets may be different with regard to the ownership of the bank. We therefore decided to estimate Model 2 with an interaction term (FOREIGN*LOANS) and removed the FOREIGN variable from the set of explanatory variables. The interaction terms were significant (at 1%) for both the cost and revenue efficiency models.²⁵ For the cost efficiency model, the coefficient for the interaction term was positive but much smaller (in absolute terms) than the (negative) coefficient for the LOANS variable. This suggests that if foreign banks have a higher loans-to-assets ratio, the decrease in cost efficiency is smaller compared to domestic banks. For the revenue efficiency model, the coefficient for the interaction term was negative but again much smaller (in absolute terms) than the (positive) coefficient for the LOANS variable. Thus, foreign banks with a higher loans-to-assets ratio were less revenue efficient than domestic banks. Therefore, ownership did not change our more general findings on the effect of LOANS on cost and revenue efficiency.

5.3 Sensitivity Analysis

To check the robustness of our results, we decided to split our sample into four years. For each year, we repeated the procedures described above, i.e., detecting and removing influential observations, testing the independence assumption, calculating bias-corrected efficiency scores, testing the separability assumption, and calculating the truncated regression as in Simar and Wilson (2007), but at the country level, only country dummies were included. This approach specifies different MCSY frontiers, and the number of observations decreased.²⁶ Therefore, some variability is expected in the results.

²⁵ As the signs and significances (even *pseudo R*²) were unchanged, we decided to conserve space and not report explicitly the results of the estimation of these models. They are available upon request.

²⁶ For the cost efficiency model, $N = 166$ for the 2005 sample, $N = 167$ for the 2006 sample, $N = 168$ for the 2007 sample, and $N = 160$ for the 2008 sample. For the revenue efficiency model, $N = 164$ for the 2005 sample, $N = 168$ for the 2006 sample, $N = 168$ for the 2007 sample, and $N = 159$ for the 2008 sample.

For the cost efficiency model, the conclusions regarding the equity-to-assets ratio remained unchanged, as it was positive and significant for each year.²⁷ The size of the bank was positively associated with cost efficiency in all years except 2008. In 2008, the FOREIGN variable was significant, while this was not the case in the other years. The result for the loans-to-assets ratio was only significant in 2007, while in 2008, the coefficient was only marginally non-significant (at the 10% significance level). The revenue efficiency results were much more similar, as at the 10% significance level, all of the bank-specific variables besides FOREIGN were significant and had the same signs as in the MCMY analysis. Overall, the results support the previous conclusions; however, this analysis also revealed that in 2008, some relationships were not as strong, possibly due to the sub-prime crisis.

6. Conclusion

Over the last 20 years, CEEC have undergone drastic economic transitions, including in their financial and banking systems. This paper studied the determinants of efficiency for commercial banks in 11 CEEC using data from 2005 to 2008. Compared to previous studies on banking in CEEC, we used a more homogeneous sample of commercial banks, using DEA to estimate both the cost and revenue efficiency of banks. Overall, the analysis of cost and revenue efficiency allowed us to identify two different banking behaviors. Such findings would be complicated to establish with cost and profit efficiency analyses alone.

Our main findings are that: i) both the size and financial capitalization of the bank are positively associated with cost and revenue efficiency; ii) foreign banks in CEEC are more cost efficient but less revenue efficient than domestic banks, suggesting different banking behaviors between foreign banks (less risky, more cost focused) and domestic banks (more risky, more revenue focused); and iii) the loans-to-assets ratio was negatively associated with cost efficiency but positively related to revenue efficiency, further stressing the two different aspects of banking behavior (cost minimization with lower loans-to-assets ratios vs. revenue maximization with higher loans-to-assets ratios).

These results have some policy implications, as bank size and financial capitalization can be regulated, and according to our results, larger and more capitalized banks had higher cost and revenue efficiency in CEEC. It seems that higher revenue efficiency is driven by a higher extent of loans, which in turn suggests higher risk. This behavior (toward revenue efficiency) is also rewarded by higher profitability. This has implications for policy makers, as taking too many risks might influence the stability of the banking system. Our empirical analysis revealed that this revenue maximization behavior was less apparent for foreign banks. One of the possible explanations might be that the risk management of foreign banks is stricter, or that the managers of foreign banks adhere more to the rules of corporate governance than the managers of domestic banks. As a consequence, foreign banks have higher standards for granting loans than domestic banks in CEE. A more general explanation might be that foreign banks adopt strategies similar to the strategies used in their “home-markets” We assume that these “home-markets” are (in the case of more developed economies) more mature and have lower growth. In such markets, cost

²⁷ Detailed results are available upon request.

efficiency might be a more natural approach than striving for revenue maximization. Finally, as domestic banks might have better knowledge of the local markets, they price risk better, which allows them to acquire higher revenue efficiency through higher exposure to loans. Future studies could address these issues in more detail.

APPENDIX A

Definition of Variables

Variable	Description
<i>Outputs and inputs</i>	
Total loans	Output variable obtained directly from Bankscope
Other earning assets	Output variable obtained directly from Bankscope
Total deposits	Input variable obtained directly from Bankscope
Total costs	Input variable constructed as the sum of interest expenses, noninterest expenses, and personnel expenses
<i>Prices of outputs and inputs</i>	
Price of total loans	Defined as the ratio of interest received on loans to total performing loans
Price of other earning assets	Defined as the ratio of noninterest income (other operating income) to other earning assets
Price of total deposits	Ratio of interest expenses on deposits to total deposits
Price of total costs	Ratio of total costs to total assets
<i>Explanatory variables (bank structure)</i>	
Average capital ratio (ETA)	Proxy for capital strength measured as a bank's equity over total assets (also termed equity-to-assets)
Return on average equity (ROAE)	Proxy for a bank's profitability
Loan risk ratio (LOANS)	Proxy for risk on loans, measured as total loans to total assets (also termed loans-to-assets)
Logarithm of total assets (LNTA)	Proxy for a bank's size measured as the natural logarithm of total assets
Foreign ownership (FOREIGN)	Proxy for a bank's ownership measured as a dummy variable with a value of 1 if more than 50% of a bank's total stock of shares is held by foreign shareholders and 0 otherwise.
<i>Explanatory variables (macroeconomic environment)</i>	
GDP per capita (GDP/C)	Used as a measure of the mean income of the population within a country based on purchasing-power-parity
Real GDP growth (Δ GDP)	Computed as the annual change in real GDP
Inflation measure (INFLATION)	Computed as the annual change in the GDP deflator with 2000 as the base year
Index of financial freedom (IFF)	Considers several aspects of government involvement in the financial system
<i>Other control variables</i>	
Hirschman-Herfindahl index (HHI)	Defined as the concentration index on total bank deposits
Country dummies	Dummy variable that takes a value of 1 if the country is one of the 11 CEEC and 0 otherwise
Year dummies	Dummy variable that takes a value of 1 if the year is one of the four analyzed years (2005–2008) and 0 otherwise

Notes: The sources of the variables are as follows: Bankscope database and authors' own calculations on banking data. GDP/C, Δ GDP, and INFLATION are obtained from the IMF and IFF from the Heritage Foundation.

APPENDIX B

Detection of Influential Observations in DEA Models—Sousa and Stošić (2005)

Following Sousa and Stošić (2005), let $\{\theta_i, i = 1, 2, \dots, n\}$ be the set of efficiency scores calculated from the original data and denote as J the set of DMUs that are removed from the sample. Then, $\{\theta_{i,J}^*, i = 1, 2, \dots, n, i \notin J\}$ is a set of efficiencies calculated for the remaining DMUs. The influence of the j -th DMU ($j \in J$) is:

$$l_j = \sqrt{\frac{\sum_{i=1, i \notin J}^n (\theta_{i,J}^* - \theta_i)^2}{n - p}} \quad (\text{B.1})$$

If $p = |J| = 1$, we need to calculate Eq. (B.1) $n-1$ times. With $p > 1$, we may be able to mitigate the masking effect (for further discussion see Sousa and Stošić, 2005, pp. 163–164). The recommended approach is to randomly select p DMUs (in one draw) and calculate Eq. (B.1). In this step, each of the removed DMUs receives the same influence. This process is repeated B times, after which we calculate the average influence for each DMU as follows:

$$\bar{l}_i = \sum_{r=1}^{B(i)} l_{i,r} / B(i) \quad (\text{B.2})$$

where $B(i)$ is the number of times that the i -th DMU was removed from the sample and $l_{i,r}$ is the r -th influence of the i -th DMU. The DMUs are then ranked according to their average influence. Now, one may either use some threshold value to indicate influential DMUs (e.g., the i -th DMU is influential if $\bar{l}_i > \tilde{l} \log n$, where \tilde{l} is the mean of \bar{l}_i) or test the equality of the efficiency distributions with and without influential observations using the Kolmogorov-Smirnov (K - S) test (see Sousa and Stošić, 2005). We selected the former as it is unclear whether using the K - S test is justified when the observations are not independent, as is clearly the case for the efficiency scores. We selected $p = \text{int}(0.15n)$ and $B = 2,500$ (or $B = 7,500$ for the MCMY frontier). However, as the choice of p and B is arbitrary and it is obvious that they may influence the rankings and leverages, we checked the robustness of our findings by trying all combinations of the following values $p \in \{0.10n, 0.15n, 0.2n\}$, $B \in \{1,000, 2,500, 5,000\}$, and $B \in \{5,000, 7,500, 10,000\}$ for the MCMY frontiers. Although some differences were found, the selection of $p = \text{int}(0.15n)$ and $B = 2,500$ (or $B = 7,500$) proved to be quite robust.

APPENDIX C

Testing the Independence Assumption

Let $1 \leq \hat{\delta}_i = \delta(\mathbf{x}_i, \mathbf{y}_i | \hat{P}) = \hat{\theta}_i^{-1}; \forall i, i=1, 2, \dots, n$ be the efficiency estimate of the i -th DMU based on the distance function δ , where \hat{P} is the production set, with \mathbf{x}_i and \mathbf{y}_i being the input and output column vectors, respectively, and $\hat{\theta}$ being the cost efficiency (for the output-oriented model, the transformation is similar, see, e.g., Simar and Wilson, 2001, p. 164). The Cartesian coordinates $(\mathbf{x}_i, \mathbf{y}_i)$ are transformed to cylindrical coordinates $(\tau_i, \boldsymbol{\eta}_i, \mathbf{y}_i)$, where $\tau_i = \sqrt{\mathbf{x}_i^T \mathbf{x}_i}$ and the j -th element of $\boldsymbol{\eta}_i$ equals $\arctan(x_{i,j+1} / x_{i,1})$ if $x_{i,1} > 0$ and $\pi / 2$ if $x_{i,1} = 0$. Note that if p is the number of inputs (and q is the number of outputs), $j = 1, 2, \dots, p-1$. Thus, for each DMU, we have $\mathbf{w}_i = [\eta_{i,1}, \dots, \eta_{i,p-1}, y_{i,1}, \dots, y_{i,q}]$. We wish to test the null hypothesis of independence between δ and \mathbf{w} , which will be denoted as $\delta \perp w$. The general idea behind the following test statistics is that the null hypothesis implies that \mathbf{w} contains no information about δ . Thus, we may write the following (Wilson, 2003):

$$H_0 : \delta \perp w \Leftrightarrow f(\delta | \mathbf{w}) = f(\delta) \quad (\text{C.1})$$

where $f(\cdot)$, $f(\cdot | \cdot)$ are the density and conditional density, respectively. We consider the following two statistics:

$$\hat{T}_{4n} \equiv \sum_{i=1}^n \left(\hat{F}_n(\hat{\delta}_i, \mathbf{w}_i) - \hat{F}_n(\hat{\delta}_i) \hat{F}_n(\mathbf{w}_i) \right)^2 \quad (\text{C.2})$$

$$\hat{T}_{5n} \equiv \max_i \left| \hat{F}_n(\hat{\delta}_i, \mathbf{w}_i) - \hat{F}_n(\hat{\delta}_i) \hat{F}_n(\mathbf{w}_i) \right| \quad (\text{C.3})$$

where $\hat{F}_n(\hat{\delta}_i)$, $\hat{F}_n(\mathbf{w}_i)$ and $\hat{F}_n(\hat{\delta}_i, \mathbf{w}_i)$ are empirical distribution functions:

$$\hat{F}_n(\hat{\delta}_i) = n^{-1} \sum_{l=1}^n I(\hat{\delta}_l \leq \hat{\delta}_i) \quad (\text{C.4})$$

$$\hat{F}_n(\mathbf{w}_i) = n^{-1} \sum_{l=1}^n I(\mathbf{w}_l \leq \mathbf{w}_i) \quad (\text{C.5})$$

$$\hat{F}_n(\hat{\delta}_i, \mathbf{w}_i) = n^{-1} \sum_{l=1}^n I(\mathbf{w}_l \leq \mathbf{w}_i) I(\hat{\delta}_l \leq \hat{\delta}_i) \quad (\text{C.6})$$

where $I(\cdot)$ is an indicator function returning 1 if the condition is true and 0 otherwise. To test the null hypothesis, Wilson (2003) recommended a standard homogeneous bootstrap algorithm:

1. Resample n efficiency estimates using the homogeneous bootstrap method of Simar and Wilson (1998), $\{\hat{\delta}_i^b, i = 1, 2, \dots, n\}$. The efficiencies are resamples from a smoothed Gaussian kernel density estimate that accounts for the boundary conditions of the efficiency estimates using the reflection method of Silverman (1986). The bandwidth in the kernel density estimation was calculated using the unbiased cross-validation method. For a more rigorous description, see the seminal works of Simar and Wilson (1998, pp. 55–56) and Wilson (2008).
2. Resample n vectors from the set $\{\mathbf{w}_i, i = 1, 2, \dots, n\}$ and obtain $\{\mathbf{w}_i^b, i = 1, 2, \dots, n\}$. Note that by using this sequence of bootstraps, we are resampling under the null hypothesis of independence.
3. Compute test statistics (2)–(3) using $\{\hat{\delta}_i^b, i = 1, 2, \dots, n\}$ and $\{\mathbf{w}_i^b, i = 1, 2, \dots, n\}$.
4. Repeat Steps 1–3 B times, acquiring the set $\{TS^b, b = 1, 2, \dots, B\}$. In our application, we used $B = 1000$.
5. Find the appropriate $1 - \alpha$ percentile or p -value.

APPENDIX D

Testing the Separability Assumption

We will illustrate the procedure for the case of the irrelevant inputs test. Let \mathbf{X} be the $n \times (p + 2)$ matrix of all inputs, which may be written using vectors as $\mathbf{X} = (\mathbf{TD}, \mathbf{TC}, \mathbf{LNTA}, \mathbf{ETA})$, with p corresponding to the number of assumed known inputs (\mathbf{TD} and \mathbf{TC}), $\mathbf{X}^p = (\mathbf{TD}, \mathbf{TC})$, and $\bar{\mathbf{Y}}$ being the matrix of price-based outputs ($n \times q$). We wish to test the null hypothesis that \mathbf{LNTA} and \mathbf{ETA} do not contribute to the production of outputs. Otherwise, the outputs depend not only on \mathbf{TD} and \mathbf{TC} , but also on \mathbf{LNTA} and \mathbf{ETA} , and the separability assumption is in question. In this context, an output-oriented model should be used, as discussed by Simar and Wilson (2001).

The test statistics are based on comparing the efficiencies obtained from the revenue efficiency model, $1 \leq \hat{\delta}_i = \delta(\mathbf{x}_i, \bar{\mathbf{y}}_i | \hat{P}) = \theta_i^{-1}; \forall i, i = 1, 2, \dots, n$ with $1 \leq \hat{\delta}_i^p = \delta(\mathbf{x}_i^p, \bar{\mathbf{y}}_i | \hat{P}^p) = (\theta_i^p)^{-1}; \forall i, i = 1, 2, \dots, n$, where \mathbf{x}_i , \mathbf{x}_i^p , and $\bar{\mathbf{y}}_i$ are the input and output column vectors, \hat{P} , \hat{P}^p are the production sets and $\hat{\theta}_i$, $\hat{\theta}_i^p$ are the revenue efficiencies. Note that for both the cost and revenue efficiency models, $0 < \hat{\theta} \leq 1$, and therefore, the test statistics are the same (for both models, we use the input-oriented specification of Simar and Wilson, 2001). Next, we define:

$$\hat{\rho}_{1,i} = \hat{\delta}_i^p / \hat{\delta}_i \geq 1 \tag{D.1}$$

$$\hat{\tau}_{1,i} = \hat{\delta}_i^p - \hat{\delta}_i \geq 0 \quad (\text{D.2})$$

Simar and Wilson (2001) proposed six test statistics, from which (for simplicity) we selected the first two:

$$\hat{\gamma}_{1,1} = \sum_{i=1}^n (\hat{\rho}_{1,i} - 1) \geq 0 \quad (\text{D.3})$$

$$\hat{\gamma}_{1,2} = \sum_{i=1}^n \left(\frac{\hat{\tau}_{1,i}}{\hat{\delta}_i} \right)^2 \geq 0 \quad (\text{D.4})$$

The critical values are obtained as follows:

1. Resample one pseudo sample of p inputs and q outputs using the heterogeneous bootstrap as in Simar and Wilson (2000, pp. 788–789). Thus, \mathbf{x}_i^{p*} and $\bar{\mathbf{y}}_i^*$ are acquired.
2. Resample one pseudo sample of potentially irrelevant inputs, uniformly and with replacement. These inputs may be denoted \mathbf{x}_i^{r*} , where r is the number of irrelevant inputs.
3. Calculate the efficiency scores $1 \leq \hat{\delta}_i^{p*} = \delta \left(\mathbf{x}_i^{p*}, \bar{\mathbf{y}}_i^* \mid \hat{P}^{p*} \right) = \left(\theta_i^{p*} \right)^{-1}$, and after stacking both of the resampled inputs, calculate the scores $1 \leq \hat{\delta}_i^* = \delta \left(\mathbf{x}_i^*, \bar{\mathbf{y}}_i^* \mid \hat{P}^* \right) = \left(\theta_i^* \right)^{-1}$. Next, calculate the test statistics. Note that because potentially irrelevant inputs were randomly resampled, we calculate the test statistics under the null hypothesis.
4. Repeat Steps 1–3 B times to obtain a set of bootstrap values of $\hat{\gamma}_{1,1}$, $\hat{\gamma}_{1,2}$. Because these procedures are computationally demanding, we use $B = 500$.
5. Compute the corresponding $1 - \alpha$ critical values or p -values.

Testing for irrelevant outputs is very similar. We define \mathbf{Y} to be the $n \times (q + 2)$ matrix of all outputs $\mathbf{Y} = (\mathbf{TL}, \mathbf{OEA}, \mathbf{LOANS}, \mathbf{ROAE})$, with q corresponding to the assumed known outputs (\mathbf{TL} and \mathbf{OEA}), $\mathbf{Y}^p = (\mathbf{TL}, \mathbf{OEA})$, and $\bar{\mathbf{X}}$ being the matrix of price-based inputs ($n \times p$). We wish to test the null hypothesis that \mathbf{LOANS} and \mathbf{ROAE} do not contribute to the production of outputs. Otherwise, the outputs depend not only on \mathbf{TL} and \mathbf{OEA} , but also on \mathbf{LOANS} and \mathbf{ROAE} . In this context, an input-oriented model should be used, as described by Simar and Wilson (2001).

APPENDIX E

Bootstrapping the Truncated Regression

1. Using the heterogeneous bootstrap algorithm, obtain the bias-corrected efficiency scores $\hat{\theta}_i$ for each bank i in the sample of n banks.
2. Use the maximum likelihood method to obtain estimates of the parameters $\hat{\beta}$ of β , $\hat{\alpha}$ of α , and $\hat{\sigma}_\varepsilon$ of σ_ε in Eq. (3) using the $m < n$ observations from a sample, where $\hat{\theta}_i < 1$.
3. Repeat Steps 3.1–3.3 B times to obtain a set of bootstrap efficiency estimates $(\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)_b, b = 1, \dots, B$.
 - 3.1. For $\forall i = 1, \dots, m$, draw ε_i^* from the distribution $N(0, \hat{\sigma}_\varepsilon^2)$, with left truncation at $(-\mathbf{Z}_i\hat{\beta})$ and right truncation at $(1-\mathbf{Z}_i\hat{\beta})$.
 - 3.2. For each $\forall i = 1, \dots, m$, compute $\theta_i^* = \hat{\alpha} + \mathbf{Z}_i\hat{\beta} + \varepsilon_i^*$.
 - 3.3. As in Step 2, we estimate parameters $\hat{\alpha}^*, \hat{\beta}^*, \hat{\sigma}_\varepsilon^*$ of the truncated regression model using the bootstrapped efficiency scores and environmental variables.
4. After $B = 3,000$ samples are drawn, calculate the p -values for each parameter using the following formula: $2\min\{\#(bp \leq 0)/B, 1 - \#(bp \leq 0)/B\}$, where bp denotes a bootstrapped parameter.

REFERENCES

- Alcober MG, Illueca M, Tortosa-Ausina E (2011): *A distributional comparison of cost and revenue efficiency for Spanish financial institutions*. XIV Applied Economics Meeting, Huelva. Available at: <http://www.alde.es/encuentros/anteriores/xiveea/trabajos/g/pdf/155.pdf>
- Andrews DF, Pregibon D (1978): Finding the outliers that matter. *Journal of the Royal Statistical Society*, 40(1):85–93.
- Banker RD, Natarajan R (2008): Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research*, 56(1):48–58.
- Benito-López B, del Moreno-Enguix MR, Solana-Ibañez J (2011): Determinants of efficiency in the provision of municipal street-cleaning and refuse collection services. *Waste Management*, 31(6):1099–1108.
- Berger AN, Humphrey DB (1997): Efficiency of financial institutions: international survey and directions for future research. *European Journal of Operational Research*, 98(2):175–212.
- Berger AN, Hasan I, Zhou M (2009): Bank ownership and efficiency in China: What will happen in the world's largest nation? *Journal of Banking and Finance*, 33(1):113–130.
- Bonin JP, Hasan I, Wachtel P (2005): Bank performance, efficiency and ownership in transition countries. *Journal of Banking and Finance*, 29(1):31–53.
- Carvalho O, Kasman A (2005): Cost efficiency in the Latin American and Caribbean banking systems. *Journal of International Financial Markets, Institutions and Money*, 15(1):55–72.
- Casu B, Molyneux P (2003): A comparative study of efficiency in European banking. *Applied Economics*, 35(17):1865–1876.
- Cazals C, Florens JP, Simar L (2002): Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1):1–25.
- Cetorelli N, Gambera M (2001): Banking market structure, financial dependence and growth: international evidence from industry data. *The Journal of Finance*, 56(2):617–648.
- Charnes A, Cooper WW, Rhodes E (1978): Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.
- Chortareas GE, Girardone C, Ventouri A (2011): Bank supervision, regulation, and efficiency: evidence from the European Union. *Journal of Financial Stability*, 8(4):292–302.
- Chronopoulos DK, Girardone C, Nankervis JC (2011): Are there any cost and profit efficiency gains in financial conglomeration? Evidence from the accession countries. *The European Journal of Finance*, 17(8):603–621.
- Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005): *An Introduction to Efficiency and Productivity Analysis*. Springer, New York.
- Cooper WW, Seiford LM, Tone K (2007): *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer, New York.
- Debreu G (1951): The coefficient of resource utilization. *Econometrica*, 19(3):273–292.
- Dinger V (2009): Do foreign-owned banks affect banking system liquidity risk? *Journal of Comparative Economics*, 37(4):647–657.
- Eisenbeis RA, Ferrier GD, Kwan SH (1999): The informativeness of stochastic frontier and programming frontier efficiency scores: cost efficiency and other measures of bank holding company performance. *Federal Reserve Bank of Atlanta, Working Paper*, no. 99–23.
- Farrell MJ (1957): The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3):253–290.
- Fethi MD, Pasiouras F (2010): Assessing bank efficiency and performance with operational research and artificial intelligence techniques: a survey. *European Journal of Operational Research*, 204(2):189–198.

- Fries S, Taci A (2005): Cost efficiency of banks in transition: evidence from 289 banks in 15 post-communist countries. *Journal of Banking and Finance*, 29(1):55–81.
- Greenbaum SI (1967): Competition and efficiency in the banking system—empirical research and its policy implications. *Journal of Political Economy*, 75(4):461–479.
- Grigorian DA, Manole V (2006): Determinants of commercial bank performance in transition: an application of Data Envelopment Analysis. *Comparative Economic Studies*, 48(3):497–522.
- Hahn FR (2007): Environmental determinants of banking efficiency in Austria. *Empirica*, 34(3): 231–245.
- Hasan I, Marton K (2003): Development and efficiency of the banking sector in a transitional economy: Hungarian experience. *Journal of Banking and Finance*, 27(12):2249–2271.
- Hauer D (2005): Explaining efficiency differences among large German and Austrian banks. *Applied Economics*, 37(9):969–980.
- Havrylychuk O (2006): Efficiency of the Polish banking industry: foreign versus domestic banks. *Journal of Banking and Finance*, 30(7):1975–1996.
- Iršová Z (2009): Measuring bank efficiency. Charles University in Prague, *Faculty of Social Sciences*, Institute of Economic Studies, *Master Thesis*.
- Iršová Z, Havránek T (2010): Measuring bank efficiency: a meta-regression analysis. *Prague Economic Papers*, 2010(4):307–328.
- Jamasb T, Pollitt M, Triebs T (2008): Productivity and efficiency of US gas transmission companies: A European regulatory perspective. *Energy Policy*, 36(9):3398–3412.
- Jemric I, Vujcic B (2002): Efficiency of banks in Croatia: A DEA approach. *Comparative Economic Studies*, 44(2–3):169–193.
- Kasman A, Yildirim C (2006): Cost and profit efficiencies in transition banking: The case of new EU members. *Applied Economics*, 38(9):1079–1090.
- Košak M, Zorić J (2011): Accounting for heterogeneity in cross-country bank efficiency comparisons. *Economics of Transition*, 19(3):473–494.
- Lehner M, Schnitzer M (2008): Entry of foreign banks and their impact on host countries. *Journal of Comparative Economics*, 36(3):430–452.
- Levin R (1997): Financial development and economic growth: views and agenda. *Journal of Economic Literature*, 35(2):688–726.
- Liu JS, Lu LYY, Lu WM, Lin BJJ (2012): Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, 41(1):3–15.
- Lozano-Vivas A, Pasiouras F (2010): The impact of nontraditional activities on the estimation of bank efficiency: International evidence. *Journal of Banking and Finance*, 34(7):1436–1449.
- Mamatkakis E, Staikouras C (2008): Bank efficiency in the new European Union member states: Is there convergence? *International Review of Financial Analysis*, 17(5):1156–1172.
- Maudos JG de, Juan F (2004): Factors explaining the interest margin in the banking sectors of the European Union. *Journal of Banking and Finance*, 28(9):2259–2281.
- McDonald J (2009): Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, 197(2):792–798.
- Mester LJ (1996): A study of bank efficiency taking into account risk-preferences. *Journal of Banking and Finance*, 20(6):1025–1045.
- Nikiel EM, Opiela TP (2002): Customer type and bank efficiency in Poland: Implications for emerging market banking. *Contemporary Economic Policy*, 20(3):255–271.
- Papke LE, Wooldridge JM (1996): Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.
- Poghosyan T, Kumbhakar SC (2010): Heterogeneity of technological regimes and banking efficiency in former socialist economies. *Journal of Productivity Analysis*, 33(1):19–31.

- Sealey CW, Lindley JT (1977): Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance*, 32(4):1251–1266.
- Silverman BW (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Simar L, Wilson PW (1998): Sensitivity analysis of efficiency scores: how to bootstrap in non-parametric frontier models. *Management Science*, 44(1):49–61.
- Simar L, Wilson PW (2000): A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6):779–802.
- Simar L, Wilson PW (2001): Testing restrictions in nonparametric efficiency models. *Communications in Statistics—Simulation and Computation*, 30(1):159–184.
- Simar L, Wilson PW (2002): Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1):115–132.
- Simar L (2003): Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.
- Simar L, Wilson PW (2007): Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64.
- Simar L, Wilson PW (2011): Two-stage DEA: Caveat emptor. *Journal of Productivity Analysis*, 36(2):205–218.
- Sousa MDCS de, Stošić B (2005): Technical efficiency of the Brazilian Municipalities: Correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, 24(2):157–181.
- Stavárek D (2006): Banking efficiency in the context of European integration. *Eastern European Economics*, 44(4):5–31.
- Tone K (2002): A strange case of the cost and allocative efficiencies in DEA. *Journal of the Operational Research Society*, 53(11):1225–1231.
- Weill L (2003): Banking efficiency in transition economies. *Economics of Transition*, 11(3):569–592.
- Wieneke A, Gries T (2011): SME performance in transition economies: The financial regulation and firm-level corruption nexus. *Journal of Comparative Economics*, 39(2):221–229.
- Wilson PW (1993): Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business and Economic Statistics*, 11(3):319–393.
- Wilson PW (2003): Testing independence in models of productive efficiency. *Journal of Productivity Analysis*, 20(3):361–390.
- Wilson PW (2008): FEAR: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences*, 42(4):247–254.
- Wilson PW (2010): *Detecting outliers in deterministic nonparametric frontier models with multiple outputs: correction*. Department of Economics, Clemson University. Available at: <http://business.clemson.edu/economic/faculty/wilson/Papers/ap-corrected.pdf>
- Yang Z, Wang X, Sun D (2010): Using the bootstrap method to detect influential DMUs in data envelopment analysis. *Annals of Operations Research*, 173(1):89–103.
- Yildirim HS, Philippatos GC (2007): Efficiency of banks: recent evidence from the transition economies of Europe, 1993–2000. *The European Journal of Finance*, 13(2):123–143.