Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP02/2021

Bank Performance Analysis

Natalya Zelenyuk and Valentin Zelenyuk

Date: January 2021

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

ISSN No. 1932 - 4398

This is a draft chapter/article. The final version will be available in due course and is to be cited as:
Zelenyuk, N. and V. Zelenyuk (2021) Bank Performance Analysis. In: N. Hashimzade and M. Thornton (Eds), Handbook of Research Methods and Applications in Empirical Microeconomics. forthcoming, Edward Elgar Publishing, Ltd.

Bank Performance Analysis

Natalya Zelenyuk* Valentin Zelenyuk †

January 12, 2021

Abstract

The goal of this chapter is to overview the current state of the art analysis of banking performance. For this, we navigate through the literature that has been prospering during the recent decades. In particular, we start with a brief discussion of the ratio analysis for measuring bank performance, which is still very popular in practice. Then we consider such popular productivity and efficiency analysis methods as data envelopment analysis (DEA) and stochastic frontier analysis (SFA). Then, we provide a brief review of other econometric methods that became leading in the recent finance literature that involve techniques of casual inference, including difference-in-differences (DD) and regression discontinuity design (RDD).

Keywords: Banking, Performance Analysis, Productivity, Efficiency, Data Envelopment Analysis, Stochastic Frontier Analysis, Econometrics, Panel Data, Causal Inference.

Acknowledgment: Both authors have contributed to all sections, and about equally overall, yet with some degree of specialization. Specifically, Natalya Zelenyuk mainly focused on banking aspects and especially on Sections 2 and 6, while Valentin Zelenyuk mainly focused on DEA and SFA methodology (especially Section 3, 4, 5). We thank many colleagues, Karligash Glass, Woo-Young Kang, Dimitris Margaritis, Bao Hoang Nguyen, Evelyn Smart, Tom Smith, Pravin Trivedi, Zhichao Wang, etc. for their fruitful feedback that helped shape this work.

*School of Business, University of Queensland, 506 Joyce Ackroyd Building (37), St Lucia, Brisbane, Qld 4072, Australia; e-mail: n.zelenyuk@uq.edu.au

†School of Economics, University of Queensland, 530 Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia; Phone:+ 61 7 3346 7054 e-mail: v.zelenyuk@uq.edu.au

1 Introduction

Banks play a major role in any economy through money creation, borrowing and lending, which fuels economic activity. Financial intermediation of banks has a direct and indirect impact on all the types of firms and their production capacities that can be restricted by or enriched with lending from banks. Banks are studied in economic literature as special firms, because of their connecting role between the financial markets and real economic growth (Sealey Jr. and Lindley (1977), Diamond and Dybvig (1983), Mester (1993), Bikker and Bos (2008)). Indeed, liquidity provided by banks to other industries is instrumental for an economy at any time, and especially during times of economic uncertainty. The importance of banks is evidenced by the attention to it from governments. During 2007 - 2009 only, the U.S. Government allocated trillions of dollars to restore liquidity creation in the economy (Bai et al. (2018), Berger and Bouwman (2009)).

Governments monitor the performance of banks and introduce various regulations aiming at the direct enhancement of the performance of banks and the indirect prevention of any liquidity deficits to develop the productive functioning of firms (Casu et al. (2004)). Besides the national levels, banking is monitored and regulated on an international level. Specifically, a key global Accord of banking regulation is documented in the Basel Rules (BIS (2020a)). The Accord was developed by the Basel Committee and directs internationally converging regulatory rules. The Basel Committee was established in 1974, in response to an increased volatility in the currency exchange markets and a failure of key banks (a prominent example of Herstatt in Germany) betting on the direction of currency pricing trends (Goodhart (2011)). The cyclicality of crisis further occurring in 1980s with the Latin American debt crisis, 1990s dot-com bubble and the 2000s Global Financial Crisis reinforced the need for the removal of the sources of competitive inequality and harmonized the regulatory requirements to bank capital, liquidity and transparency (BIS (2020b)). The most recent regulatory accords are identically standardized and implemented almost in every country in the world and are based on the advanced performance methods balancing risk with return.

The goal of this chapter is to present practical tools that have been extensively applied to performance analysis of banks. The toolset in this chapter covers ratio analysis, efficiency techniques of Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) as well as it outlines other econometric methods, including causal inference that recently became very popular in financial literature, comprising of difference-in-differences and regression discontinuity design methods. We believe that our chapter will be valuable for academics, including PhD students, studying banking as well as banking industry experts and professionals engaged in research and consulting, regulation and prudential monitoring and analysis of banks.

1.1 DEA and SFA

DEA has its roots in economic theory models, mainly known as *Activity Analysis Models* (AAM), developed by Koopmans (1951), Debreu (1951), Shephard (1953), which were in turn influenced by the works of Leontief (1925), Von Neumann (1945), to mention just a few. *Inter alia*, these and other important works led to the seminal paper of Farrell (1957), who tailored the AAMs into the foundation for measuring what he referred to as 'productive efficiency'. Besides a few papers, his work was largely overlooked for two decades, until two streams of research basically exploded from his work. One stream was largely due to the influential research of Charnes et al. (1978), where Farrell's approach was generalized and empowered with linear programming theory and the practice to implement it, as well as being branded as DEA. The second stream was largely due to the seminal works of Aigner et al. (1977) and Meeusen and van den Broeck (1977), who at about the same time proposed SFA, building on the research of Farrell (1957) among others.

Over the last four decades, both of these streams, DEA and SFA, have developed into very rich streams of literature, with many branches, often interconnected and mutually developing and spanning several fields of research: Economics and Econometrics, Statistics (including Data Science & Business Analytics), Operations Research/Management Science, etc. A key, and in some respects more important, part of the literature is about the applications of these methods virtually to any sector of the economy. Among the most popular applications are those to the banking industry, which we briefly overview below and give references for further details.

Among the first papers using DEA in banking were Sherman and Gold (1985), English et al. (1993), Fukuyama (1993), which were followed by many others. SFA came out with the works of Ferrier and Lovell (1990), Hughes and Mester (1998), which were among the first applying SFA in banking, followed by many interesting works thereafter. Now, both literature streams are massive, e.g. Google Scholar search (in April 2020) yielded about 58,400 results for 'DEA and Banking' and 14,000 for 'SFA and Banking', while a more general Google search for 'DEA' and 'SFA' yielded about 6,660,000 and 2,000,000 results respectively. Obviously, it is practically impossible to review even 10% of the published papers. Our goal is more modest to overview the major methods and to mention a few popular papers that we think represent valuable examples, as well as to refer to previous reviews where more examples could be found.

1.2 Causal Inference Methods

Causal methods are rooted from the randomized experiments of Splawa-Neyman et al. (1990, 1923)[1] and Fisher (1925), where both authors introduced a design of experiments around the testing of treatment effects. These seminal works and their later developments into agricultural, biological and medical research lead to an application of the difference (in differences) methods in empirical economics (where fundamental works include Ashenfelter (1978) and Ashenfelter and Card (1985)) and finance (Harris (1966) and Trezevant (1992)).

The recent prominent papers applying DD in banking include Anginer et al. (2018) in an application to corporate governance, Buchak et al. (2018) in an application to 'fintech' lending and Duchin and Sosyura (2014) in studies of government bailouts of banks, among many others.

Another popular method that is identifying causality between the two variables is regression discontinuity design. It gained popularity from the seminal work of Thistlethwaite and Campbell (1960) and received attention in the economic and finance literature due to the minimal assumptions involved in modeling. Hahn et al. (2001) and Imbens and Lemieux

[1]This work was translated and edited in 1990 by D. M. Dabrowska and T. P. Speed from the Polish original that appeared in *Roczniki Nauk Rolniczych Tom X (1923) 1-51 (Annals of Agricultural Sciences).*

(2007) provide theoretical and empirical guidelines on RDD. In banking research, RDD applications are relatively new, a few prominent papers include the works by Bonner and Eijffinger (2015) on the impact of regulatory liquidity requirements on the demand for long-term lending, Liberman (2016) on the effects of renegotiations on credit card lending, and Berg (2018) on the effects of precautionary savings in the transmission of credit supply shocks.

1.3 Chapter Structure

Our chapter is organized as follows. Section 2 starts with the review of the development of the performance analysis in banking from the simple methods based on financial ratios to explaining inefficiency with more advanced methods (Data Envelopment Analysis, Stochastic Frontier Analysis, causal inference), Section 3 presents the basics of the productivity and efficiency theory, Section 4 reviews Data Envelopment Analysis (DEA), Section 5 reviews the Stochastic Frontier Analysis (SFA), Section 6 presents an overview of other econometric methods based on the recent developments in the causal inference analysis of performance, and Section 7 concludes.

2 Evolution of Performance Analysis of Banks

In this section we present how the methods of performance analysis of banks developed from very straightforward managerial techniques of financial statements analysis to advanced linear programming methods enabling the explanation of the inefficiency of banks.

2.1 Analysis of Financial Ratios

Traditionally, analysis of bank performance starts with an analysis of financial statements, mostly statement of a financial position and a comprehensive income statement (Casu et al. (2006)). The simplest measures of performance would include the assets of a bank and the income that was generated with these assets during a financial year. Basically, all the others, including very complex methods of performance analysis in banking, often use these measures as building components or generalize over them.

The analysis with financial ratios is based on combining two (absolute) financial measures into a ratio. For example, net income over assets would signal about how much net income a bank has generated per dollar of assets. In efficiency analysis literature, it is an example of the 'return-to-dollar' measure, advocated by Georgescu-Roegen (1951) and advanced further in Färe et al. (2019).

Bank counter-parties, including shareholders and regulators, are often interested in the performance analysis of banks with the financial ratios. Financial ratios are arithmetically simple and usually very intuitive. Various management information systems are designed around financial ratios and the breakdown of the key ratios into driving components. One of the ways of the ratio decomposition follows the famous DuPont Identity, where the interesting ratio for shareholders, return on equity, is decomposed as a product of the two other ratios—the return on assets and the equity multiplier (Ross et al. (2017)). Each of these two ratios could be further decomposed into the component-driving ratios. For example, the return on assets is a product of the profit margin and the total assets turnover. This way, one can describe the drivers of the return on equity (measured by net income/total equity) with operating efficiency (measured by net income/sales), asset use efficiency (sales/assets) and financial leverage (assets/total equity) (Ross et al. (2017)). The most detailed management information systems include decision trees that relate the top 'all bank performance' return on equity to the sales efficiency of a loan officer in a single branch.

In essence, the key financial performance ratios of interest to any bank and bank stakeholders will include the return on assets (ROA), return on equity (ROE), non-performing loan ratios and capital ratios (Casu et al. (2006)). Interestingly, after the Global Financial Crisis banks increasingly applied risk-adjusted measures of performance (BIS (2015)), including risk-adjusted return (Bikker and Bos (2008)) on capital (RAROC), defined as

$$RAROC = \frac{\text{Risk-adjusted Net Income}}{\text{Economic Capital}} \tag{1}$$

return on risk-adjusted capital (RORAC), defined as

$$RORAC = \frac{\text{Expected Net Income}}{\text{Allocated Economic Capital}} \tag{2}$$

and risk-adjusted return on risk-adjusted capital (RARORAC), defined as

$$RARORAC = \frac{\text{Economic Value Added}}{\text{Allocated Economic Capital}}. \tag{3}$$

Meanwhile, the risk-adjusted return measures the net income a business has earned in relation to a risk in a given time frame. With accounting measures it could be calculated as

$$\begin{aligned}\text{Risk-adjusted return} =& \text{Revenue} - \text{Expenses} - (\text{Expected Loss}) \\ &+ (\text{Net Interest Margin of Capital Change}).\end{aligned} \tag{4}$$

As a result, this formula ensures that a business with a lower risk and a constant level of return will have a higher risk-adjusted return. The denominator of the RAROC ratio, economic capital, attributes bank capital securing the effects of risk-taking by business. It is often parametrized to the level of capital a bank needs to handle unexpected losses during a particular time period within a decided confidence interval. For better management economic capital is calculated at a total bank level and at a level of divisions. Building blocks of economic capital vary by type of risk. For example, credit risk modeling would require an estimation of probability of default, loss given default and exposure at default (Roncalli (2020)).

The economic value added (EVA), in a numerator of the RARORAC, presents a measure of economic profit as an incremental difference between a rate of return and the cost of capital. It is calculated as a net income in a difference to an invested capital times the cost of capital. EVA is presented in absolute measures and is also calculated at the level of a bank and its divisions (BIS (2015)).

Alternatively, risk-adjusted return is measured with the asset pricing approach, where RARORAC is given by

$$RARORAC = \frac{[(r_p - r_f) - \beta_p(r_m - r_f)]\mathcal{I}_0}{\text{Economic Capital}}\,, \tag{5}$$

where r_p is return on a portfolio of assets

r_f is a return on a risk–free asset, i.e., government bond

r_m is the market return

β_p is a systematic risk of the portfolio

$\mathcal{I}_0$ is an investment in a project at time = 0, at the beginning of the project. It is a utilized capital at risk (BIS (2015)).

In the denominator, the Economic Capital, or allocated capital at risk, is a Value at Risk (VaR), which is the measure of a loss on a portfolio over a fixed time period (Roncalli (2020)).

The difference in the calculation of the two measures of RARORAC with an economic value added and with an excess return is driven by the availability of accounting versus market data applied in the calculations. For example the value added versus the loss on the loan portfolio is presented with an accounting data, whereas the return on a bond portfolio is likely to be measured with the market pricing data.

For regulatory and investment purposes, bank performance ratios are combined into the CAMELS system. The most recent governmental capital support of the banks during the GFC in the U.S. was grounded on a review of the bank applicants according to the CAMELS rating system (Duchin and Sosyura (2014)). CAMELS stands for Capital Adequacy, Asset Quality, Management, Earnings, Liquidity and Sensitivity to Market Ratios (Federal Reserve (2020)). Table 1 summarizes the key regulatory requirements corresponding to each of the CAMELS criteria and the ratios applied in research to represent these criteria.

Each of the rating measures in CAMELS is focused on balancing returns with the risks. To start with returns, every bank is aiming at increasing returns to shareholders by maximizing revenues and minimizing cost, that is by improving the efficiency of a bank (Bikker and Bos (2008)). The key to efficiency is a technology that transfers inputs (cost-related contributors to the production process of a bank) into outputs (revenue-related contributors to the production process of a bank).

In contrast to productivity that measures the output produced relative to the input, efficiency measures the difference between the actual/observed and optimal/unobserved input-output result. Thus, the challenge of estimating efficiency is in the observability of both technology and optimal efficiency. The roadmap to overcoming these challenges is discussed in the next sections.

2.2 Variables for Modeling Production Process in Banking

A critical step in the analysis of productivity and efficiency of any system is defining the inputs and outputs of the production process of this system. A conceptual framework for the economics of production analysis of banks was proposed in a prominent work by Sealey

Table 1: CAMELS Rating System, Corresponding Regulatory Expectations and Ratios

Rating Component	Uniform Regulatory Expectation from https://www.fdic.gov	Corresponding Financial Ratio
Capital Adequacy	Level of capital commensurate with the degree of credit, market and other on and off-balance sheet risks.	Tier 1 risk-based capital ratio = = (Tier 1 capital)/(risk-weighted assets) (Duchin and Sosyura (2014)). Common equity ratio = (common equity)/(total assets) (Anginer et al. (2018)).
Asset Quality	The adequacy of allowance for credit risk, including loan and other investment portfolios losses.	Non-performing loans ratio = = (loans past due 90 days or more and non-accruals)/(total loans) (Acharya and Mora (2015)).
Management	Sound management practices to ensure balanced bank performance and risk profile.	Return on assets = = (net income)/(total assets) (Martinez-Peria and Schmukler (2001)).
Earnings	Strong developing quantity and quality of earnings adequate to the risks.	Return on Equity = = (net income)/(total equity capital) (Duchin and Sosyura (2014)).
Liquidity	Adequate liquidity position compared to funding needs, availability of assets convertible to cash, access to money markets and sources of funding.	Cash ratio = cash/(total assets) (Assaf et al. (2019)).
Sensitivity to Market Risk	The degree to which changes in interest rates, foreign exchange rates, equity prices, or commodity prices can adversely affect the earnings or an economic capital of a bank. The nature and complexity of market risk exposure due to trading and foreign operations.	Sensitivity to interest rate risk = = (short term assets — short-term liabilities)/(earning assets) (Duchin and Sosyura (2014), Berger and Roman (2017)).

and Lindley (1977) defining inputs and outputs of banks in their production, installing the economics foundation for both DEA and SFA. Since then, three major approaches prevailed in the literature for modeling the technological process of banks: production, value-added and intermediation. Many studies, while arriving at the similar, or different, conclusions on the same or similar data differ by how they apply inputs and outputs to describe the technological role of banks in modeling with both DEA and SFA. Fruitful reviews on this and other aspects can be found in Berger and Humphrey (1997), Fethi and Pasiouras (2010) and Paradi and Zhu (2013), to mention just a few.

Basically, in the production approach, the role of a bank is similar to an industry decision making unit (DMU), that is labor and capital are needed to produce goods, including customer accounts, loans and securities. This approach is mostly applied in studies of the bank level and branch level efficiency rather than in the studies about an industry efficiency. For application examples, see Berger et al. (1997), Camanho and Dyson (2005), Kenjegalieva et al. (2009a).

In the value-added approach, the role of a bank is to create income with a difference between earnings from outputs and cost from inputs. This approach is particularly relevant for diversified activities, extending to insurance and 'bancassurance'. Application examples include Eling and Luhnen (2010) and Leverty and Grace (2010). Recently, Humphrey (2020) extended this approach to creating value for bank customers, where outputs are measured with value creation with both assets and liabilities of a bank, i.e., loans and securities are redefined into adding value flows that are important for users. In this model, inputs are deposits, labor and capital.

In the most commonly applied intermediation approach, the role of a bank is to transform savings (mostly deposits) into investments (mostly loans). The bank is viewed as a DMU which collects deposits with labor and capital and produces loans and other earning assets (Sealey Jr. and Lindley (1977)). Application examples of this approach include analysis at the bank level and can be found in Aly et al. (1990), Isik and Hassan (2002), Casu et al. (2013), Almanidis et al. (2019), Mamatzakis et al. (2019), to mention a few.

Implementation of inputs and outputs to formally model the banking production process is described in Section 3.

2.3 Variables for Explaining Inefficiency in Banking

One of the major aspects of interest in the analysis of productivity and efficiency of banks is modeling the inefficiency on other covariates, or explanatory variables. An established econometric strategy here is based on choosing the explanatory variables grounded on theoretical, or practical, considerations of the relevance of the covariates to the analyzed inefficiency. In a prominent methodological paper, Simar and Wilson (2007) illustrate how the algorithm of their model can work with banking data by including covariates of the size of a bank, product diversity and characteristics of the environment where a bank operates (e.g., see Aly et al. (1990), Assaf et al. (2019)). While applying the similar conceptual variable selection, the proxies for each of the concepts differ by studies and in this section we briefly overview some of the most typical choices.

Size of a bank. The common explanatory covariate that appears in many (if not all) studies of the performance analysis literature is size, due to beliefs that larger banks may utilize scale effects better. Larger banks often have greater diversification capacities (Assaf et al. (2019)). The common specifications of size include the natural logarithm of total assets (Simar and Wilson (2007)), total deposits/loans and number of bank branches (Aly et al. (1990)). By a scale effect, a common belief is that size would have a positive impact on bank efficiency. For example, Berger and Mester (1997) found a positive impact of bank size on cost efficiency and a negative impact of a bank size on profit efficiency. Separately, Aly et al. (1990) found a positive effect of size on both technical and allocative efficiency of a bank. In a more recent study, Almanidis et al. (2019) found that two top-tier large banking groups were more efficient than four smaller sized banking groups.

The banks are often differentiated by size because larger banks require more regulatory attention based on their systemic importance. Systemically important banks are considered to be "too big to fail" (Bikker and Bos (2008)). In the event of a stress, the government would need to bail these banks out because the failure of a big bank could have too big risk to the real sector in the economy (Adrian and Brunnermeier (2016)). One of the approaches to studying the impact of the size of a bank on its performance is by assigning a dummy variable of one to a larger and systemically important bank (Acharya and Mora (2015)) and zero to the rest of the banks. Most of the regulatory rules link the systemic importance of a bank to the size of a bank. The Dodd-Frank Act in the U.S. defines a banking organization as systemically important when its total assets exceed $50-$250 billion (Federal Reserve (2020)).

The primary global regulatory standard establisher, the Basel Committee on Banking Supervision, added more criteria to the definition of systemic importance. In particular, the Basel criteria also include complexity, interconnectedness, substitutability and the global significance of a banking organization (BCBS (2019)). The banks that meet the Basel criteria on significance are included in the annual lists of Global Systemically Important Banks published by the Financial Stability Board in the European Union. The list comprises of Bank of China, BNP Paribas, Citigroup, Commerzbank, Erste, Lloyds, Mitsubishi Group, to mention a few (EBA (2020)).

Many countries adopted (sometimes with modifications) the Basel criteria to their environments. For example, in Australia, Australian Prudential Regulator followed the Basel Guidelines and IMF consultations to apply all four criteria to the definition of domestic systemically important banks and has identified the four biggest banks as systemically important. Total resident assets of the smallest of these four banks, ANZ, exceed AUD 400 billion (APRA (2013)), about the size of the GDP of Singapore.

Product diversity presents the proportion of the firm revenue generated with the products that a bank offers to its customers. Greater diversification is usually preferred for a bank for *risk* reduction purposes. For example, Aly et al. (1990) and Simar and Wilson (2007) apply an index representing zero for single-product lending banks that increases with the number of products added.

Another common approach to presenting diversity is with a dollar value of a particular product in a dollar value of total assets. Isik and Hassan (2002) apply a share of loans to total assets as explanatory variables and find a positive association between the share of loans in assets and efficiency. A similar result was found in a classic study by Berger and Mester (1997), suggesting that a bank loan product could be higher valued than a bank security. More recently, Assaf et al. (2019) applied commercial real estate loans in total assets as a control variable to explain which banks perform better during the crisis and found that the higher this ratio is, the lower the risk of bank failure during the crisis. The proportion of loans in the other assets and what types of loans are to be issued by a bank comprise the strategic questions for bank management.

Risk, as one of the most important financial concepts (along with return), is applied in numerous studies (Assaf et al. (2019), Berger et al. (2010), Altunbas et al. (2007)) to explain bank inefficiency. Risk is usually proxied with the two variables: capital ratio and non–performing loans (NPL) ratio. In the simplest form, capital ratio as the share of equity to total assets, represents a bank's ability to absorb losses for 'safety and soundness' of the bank (BIS (2020a)). It is expected that higher capital ratio would signal better bank management (Assaf et al. (2019)) and better performance (Berger and Bouwman (2013)).

The impact of capitalization on efficiency is presented with mixed evidence. Conditional on a model specification, Mamatzakis et al. (2019) found both positive and negative associations between capitalization and efficiency for a sample of Japanese banks. In the study of 10 European banking systems, Lozano-Vivas et al. (2002) found that capital ratio is important for explaining efficiency differences for banks in the U.K. and France. In a recent study of 15 European countries, Altunbas et al. (2007) found a positive association between a change in capital and bank inefficiency. Overall, the majority of reviewed papers signalled a positive association between capitalization and efficiency. An explanation for a strength of the capitalization - the efficiency link could be an avenue for future research.

NPLs are often presented as a credit risk proxy. Regulators often require banks to hold higher levels of equity capital when the banks have higher than average NPL ratios. The risk arising from NPL has accumulative features. Higher current NPL ratios constrain loan production in the future periods (Fukuyama and Weber (2015)). The association between the level of NPLs and efficiency is found to be positive. Efficient banks take a higher risk (Altunbas et al. (2007)). Yet, when the NPL ratio is applied as a proxy for management (better managed banks have lower NPLs (Berger and DeYoung (1997)), an association between NPLs and efficiency appears to be negative (Koutsomanoli-Filippaki et al. (2009)). Although, risk proxies have been found to have an association with efficiency, a number of more recent studies model NPLs as an undesirable output that is reducing the value of a desirable net loan output (e.g., see related discussion in Pham and Zelenyuk (2019)).

Local economic conditions are an important determinant of inefficiency when banking systems are compared across regions. Aly et al. (1990) find a positive effect of urbanization on the efficiency of US banks. In a European cross-country study, Lozano-Vivas et al. (2002) included income per capita and salary per capita to show a positive association between the higher levels of these environmental variables and bank efficiency. Other common variables applied in international bank performance studies include the level of real GDP growth (Leary (2009)) and inflation.

Ownership represents a major part of the corporate governance puzzle, in particular, a discussion of its combination with management (Shleifer and Vishny (1997)). Ownership is usually differentiated into two categories to explain efficiency: private versus public (Berger et al. (2010) and Sturm and Williams (2004)) and foreign versus domestic (Berger et al. (2010), Sturm and Williams (2004), Berger et al. (2000)). These two types of ownership are often included in one study because foreign ownership inflow into a country often follows with a degree of deregulations and could be accompanied by privatizations.

For instance, Berger et al. (2010) document that privatization in China followed by increased foreign ownership had a positive effect on bank efficiency. In an earlier study, Berger et al. (2000) found that domestic banks were more efficient than foreign-owned banks in four countries Germany, France, Spain and the U.K. In contrast, Sturm and Williams (2004) found that foreign banks were more efficient than domestic banks in Australia, especially after complex de-regulation in the 1980s that included removal of the interest rate controls and loan quality controls.

More recently, Berger et al. (2010) suggest that the effect of ownership on efficiency differs by country due to the stages of development of the country. In particular, they discuss that foreign-owned banks are more efficient in developing countries and domestic banks are more efficient in developed countries. In this vein, Koutsomanoli-Filippaki et al. (2009) find that foreign ownership brought positive efficiency change to all the banks in Central and Eastern Europe, including domestic private and state-owned banks. On the other hand, Zelenyuk and Zelenyuk (2014) found that the efficiency of foreign-owned banks was insignificantly different from the efficiency of the domestic banks when a foreign bank owns a domestic bank partially. Foreign banks only exceed domestic banks by efficiency when foreign ownership comprises 100% of a bank.

Overall, the conclusions about the effect of ownership on efficiency are mixed, perhaps because ownership status is evolving and the agency problems associated with ownership and management are avenues that are being resolved dynamically, which might be challenging to capture with modeling.

In this sub-section, we overview the most popular ‘environmental variables' believed to be associated with the efficiency of banks. Any particular variable is research-question specific and is often limited by the data availability. The pool of explanatory variables could be extended with control variables applied in causal inference analysis that we discuss further in sub-section 2.4.

Implementation of the covariates in the formal efficiency modeling is discussed in sections 4 and 5.

2.4 Variables in Analysis of Causal Effects in Banking

In studies of causal effects, the emphasis is on a small number of parameters, often only one identifying parameter (Cameron and Trivedi (2010)). A distinction is made between an outcome variable and a treatment/forcing variable. For example, the bank interest margin could be the outcome variable, while the regulatory liquidity requirement can be thought of as a treatment variable (e.g., see Bonner and Eijffinger (2015)). The rest of the variables in the estimation are often controls.

While a choice of variables for a causal treatment effect is a challenging researcher's task that can be novel, the set of control variables is often standard. Similar to the environmental variables in sub-section 2.3, controls typically include the size of a bank and the managerial ability of balancing risk with return, reflected with the CAMELS ratios presented in Table 1.

In particular, Column (1) of Table 1 names the CAMELS component, Column (2) specifies a uniform regulatory expectation about each of the components and Column (3) defines the corresponding financial ratios. While interpreting the financial ratios, Column (3) also provides examples of how these ratios were applied as controls in the banking studies. The most comprehensive approach to controls includes both size and all of the CAMELS variables (Duchin and Sosyura (2014), Eichler et al. (2011)). The proxies for CAMELS may vary.

In a prominent study on market discipline, Martinez-Peria and Schmukler (2001) apply a return on assets (ROA) to control for both management and earnings. ROA presents an amount of net income generated with the bank assets and reflects the managerial ability of creating the former with the latter. At the same time, from an accounting point of view, income reflects the difference between earnings and cost, thereby, the ROA ratio also signals earnings quantity and quality. To further control for any heterogeneity between the banks, a common practice is to include bank fixed effects that capture all the remaining differences between the banks that are time-invariant during the study period (Duchin and Sosyura (2014)). In recent work by Boubakri et al. (2020) about the post-privatization state ownership and bank risk taking, the authors control for bank fixed effects as well as (and similarly to discussed variables in sub-section 2.3) for environmental variables of competition within the banking industry, information sharing with the private/public loan registries and creditor rights.

Due to the recent global convergence of banking regulations to uniform standards, environmental variables that describe legal and competitive local conditions gain particular significance in the cross-sectional studies (Bonner and Eijffinger (2015)).

3 Productivity and Efficiency Theory for Banking

Before estimating efficiency or productivity of banks, it is important to sketch a theoretical model that represents the essence of their production activities. For this, let $x = (x_1, ..., x_N)' \in \Re^N_+$ represent the vector of inputs that a bank, its branch or, more generically its decision making unit (DMU) uses to produce outputs, denoted by vector $y = (y_1, ..., y_M)' \in \Re^M_+$. As discussed above, in sub-section 2.2., in recent efficiency studies on banking, researchers typically proxy inputs with capital and labor while for the outputs they apply loans and other earning assets, although this may vary depending on the goals of the study.

3.1 Technology Characterization

In reality, each bank, or DMU, may use its own specific technology to produce its outputs from various inputs. A natural starting point is a simple case when technology involves only one output—then the technology (and its frontier) of a bank can be characterized with what economists call a *production function*. In general terms, a production function, call it ψ, is defined as a function yielding the maximal output producible with the knowledge and level of input (x) available at time t and conditions (Z) faced by a bank i at time t. More formally:

$$\psi_{it}(x|Z_{it}) \equiv \max_x \{y : y \text{ is producible from } x \text{ at time } t \text{ with conditions } Z_{it}\}$$

In practice, of course, banks typically produce more than one output, as a result, a more general characterization is needed. This can be done with what production economists call *technology set*, which we denote as Ψ, and define as

$$\Psi_{it}(Z_{it}) \equiv \{(x, y) : y \text{ is producible from } x \text{ at time } t \text{ with conditions } Z_{it}\}. \quad (6)$$

To compare different banks in terms of performance, one has to define a common benchmark. Often, researchers take what can be called as an egalitarian approach, in the sense that all DMUs are measured, or benchmarked, with respect to the same frontier in a given period, t, sometimes referred to as the observed 'grand frontier', or the unconditional frontier, i.e.,

$$\Psi_\tau = \cup_i \Psi_{it}(Z_{it}|t = \tau) \tag{7}$$

where the union is taken over all the possibilities available for all i in a specific time period $t = \tau$. In a sense, this is similar to some athletic competitions: e.g., everyone runs the same distance regardless of the actual technologies they used for developing their training skills for running.

In further discussions we will focus on this unconditional frontier Ψ_t but for simplicity of notation also dropping the subscript t, unless it is needed.[2] We will also assume that the technology meets the requirements of standard regularity conditions or axioms of production theory (e.g., see Sickles and Zelenyuk (2019) for more details).

Upon estimating such an unconditional frontier, a researcher may then try to analyze the association of the resulting efficiency scores (with respect to the unconditional frontier) and Z_{it} and, possibly, t. This would be the so-called two-stage efficiency analysis, popular in DEA literature for many years and more recently revitalized due to the key work of Simar and Wilson (2007), although also see the caveats discussed in Simar and Wilson (2011) and related discussion in Sickles and Zelenyuk (2019, Ch. 10). Ideally, a one-stage approach, where association of efficiency with other variables is modeled explicitly when estimating the frontier, is preferred whenever it is possible. This is particularly true in the SFA approach, as clarified by Wang and Schmidt (2002), as we will discuss in Section 5.

3.2 Relativity

Efficiency is a relative and normative concept — it always depends on the selected criterion that defines the benchmark of comparison and answers the question: 'Relative to What?' Indeed, it could very well be that a bank is very or even 100% efficient relative to one criterion, while very inefficient relative to another, or perhaps many other criteria. It is therefore imperative to clarify which criterion is selected and motivate why it was selected.

For example, a natural criterion is the frontier of the technology set, yet there are different (and not always equivalent) definitions of the frontier. A popular definition is the efficient frontier according to the Pareto-Koopmans criterion, or optimality principle, that says: (x, y) is *Pareto-Koopmans efficient* for technology Ψ if and only if with such technology it is infeasible to increase any outputs without decreasing some other outputs or increasing some of the inputs, and also infeasible to decrease any of the inputs without increasing some other

[2] It is also worth noting that when having 'too few' observations in a time period, researchers consider pooling over several time periods, i.e.,

$$\Psi = \cup_{t\in\{1,...,T'\}} \Psi_t = \cup_{t\in\{1,...,T'\}} \cup_i \Psi_{it}(Z_{i\tau}|\tau = t) \tag{8}$$

where the union is taken over all the possibilities available for all i in a specific time range $\{1, ..., T'\}$, which can be all the sample or part of the sample with the moving window approach. A caveat of this approach is that it presumes (or ignores) the technological change that may have happened over that period and, therefore, should be done with caution and relevant justifications.

inputs or decreasing some of the outputs in y. Thus, mathematically, the *Pareto-Koopmans efficient subset of the technology frontier* Ψ can be defined as

$$\mathit{eff}\,\partial\Psi \equiv \{(x,y)\ :\ (x,y)\in\Psi,\ (x^0,y^0)\notin\Psi, \forall(-x^0,y^0)\geq(-x,y)\} \tag{9}$$

Another relevant criterion is the level where the optimal scale of using resources is reached, sometimes referred to as the socially optimal level.

Yet another relevant criterion, and usually the ultimate benchmark for measuring the performance of banks, is the maximal profit feasible for a technology Ψ and some output and input prices, which can be formally stated as

$$\pi := \max_{x,y}\{p(y)y - w(x)x\ :(x,y)\in\Psi\} \tag{10}$$

where $p(y)$ is the vector of inverse demand functions for each output and $w(x)$ is the vector of inverse supply functions for each input.

From economic theory we know that the full efficiency is reached with perfect competition and this implies a relevant benchmark for comparison. Specifically, perfect competition will imply that the prices are exogenous to any individual firm and so the optimization criterion (10) simplifies to

$$\pi(p,w) = \max_{x,y}\{py - wx\ :\ (x,y)\in\Psi\} \tag{11}$$

By accepting this benchmark, various profit efficiency measures can be constructed and, in fact, many were offered in the literature. Most recently, Färe et al. (2019) proposed a very general profit efficiency measure that unified many other efficiency measures in the literature. Färe et al. (2019) also showed that the Farrell measures of technical efficiency, which are the most popular in practice, are components, or special cases, of their general profit efficiency measure. In the next section we briefly discuss some of them, while more can be found in Färe et al. (2019) and Sickles and Zelenyuk (2019).

3.3 Popular Efficiency Measures

By far the most popular efficiency measures in general, and in the context of banking in particular, are the Farrell measures of technical efficiency and, to save space, we will mainly focus on them.

Specifically, the Farrell input oriented technical efficiency can be defined as

$$ITE(x,y) \equiv \frac{1}{IDF(x,y)},\ (x,y)\in\Psi$$

where $IDF(x,y)$ is the input oriented Shephard distance function, $IDF : \Re^N_+ \times \Re^M_+ \to \Re_+ \cup \{+\infty\}$, defined as

$$IDF(x,y) \equiv \sup\{\theta > 0 : (x/\theta, y)\in\Psi\}.$$

Meanwhile, the Farrell output oriented technical efficiency can be defined as

$$OTE(x,y) \equiv \frac{1}{ODF(x,y)},\ (x,y)\in\Psi$$

where $ODF(x, y)$ is the output oriented Shephard distance function, $ODF : \Re^N_+ \times \Re^M_+ \rightarrow \Re_+ \cup \{+\infty\}$, defined as

$$ODF(x, y) \equiv \inf\{\lambda > 0 : (x, y/\lambda) \in \Psi\}.$$

These distance functions are multi-output generalizations of the notion of the production function. Indeed, note that for a one-output case, we have

$$ODF(x, y) = \frac{y}{\psi(x)}$$

where $\psi(x)$ is the production function. This special case reveals that the ODF is a natural measure of technical efficiency that relates the actual output to the potential (or maximal) output.

The Farrell technical efficiency measures described above are radial measures, in the sense that they measure inefficiency in a radial way, i.e., along the ray from the origin and through the point of interest all the way to the frontier in either input or output direction while holding, respectively, the output or input vectors fixed. This means that all the inputs (outputs) are contracted (expanded) by the same proportion, while keeping the outputs (inputs) and technology fixed.

Finally, another very general efficiency measure and a complete characterization of technology is based on the *directional distance function,* $DDF_d : \Re^N_+ \times \Re^M_+ \rightarrow \Re \cup \{+\infty\}$, defined as

$$DDF_d(x, y) \equiv \sup_\theta\{\theta \in \Re \ : \ (x, y) + \theta d \in \Psi\}, \tag{12}$$

where $d = (-d_x, d_y) \in \Re^N_- \times \Re^M_+$ is the direction that defines the orientation of measurement.[3] By choosing different directions, many efficiency measures can be derived from this function.

Under the standard regularity conditions, these distance functions (and hence the Farrell technical efficiency measures) have many desirable mathematical properties and, in particular, provide complete characterizations of technology Ψ, in the sense that

$$1/IDF(x, y) \in (0, 1] \iff ODF(x, y) \in (0, 1] \iff DDF_d(x, y) \geq 0 \iff (x, y) \in \Psi$$

From duality theory in economics, we also know that under certain conditions, the technology set can be characterized by cost, revenue and profit functions, which can be used to define the dual efficiency measures, cost efficiency, revenue efficiency and profit efficiency, which can be then decomposed into the technical efficiency and various allocative efficiencies (e.g., see Sickles and Zelenyuk (2019, Chapter 3) for details).

4 Envelopment Estimators

For several decades DEA has been a standard and very popular approach in the toolbox for performance analysis in general and for banks in particular. The literature on DEA is vast and growing. Among the first was the work of Sherman and Gold (1985), who evaluated

[3]The ideas of this measure go back to as early as 1940s, and more thoroughly developed by Chambers et al. (1996, 1998), to mention a few.

the operating efficiency of bank branches using early DEA models. This work was followed by three now seminal studies that appeared in top econometric journals: Aly et al. (1990), Charnes et al. (1990) and Ferrier and Lovell (1990). Shortly after appeared the interesting, and now classic, works of English et al. (1993) and Wheelock and Wilson (2000). While all ofthese works were focusing on US banking, besides explaining particular research questions, they helped in popularizing the novel at that time methods, and eventually they found applications for many (if not all) other countries globally.

For example, Du et al. (2018) focused on the performance of banks in China, Fukuyama (1993); Fukuyama and Weber (2015) in Japan, Casu et al. (2013) in India, Simper et al. (2017) in Korea, Curi et al. (2013, 2015) in Luxembourg, Camanho and Dyson (2005) in Portugal, etc. Meanwhile, a few studies focused on a set of countries, e.g., Casu et al. (2004), Koutsomanoli-Filippaki et al. (2009), Kenjegalieva et al. (2009a) and Lozano-Vivas et al. (2002) analyzed sub-sets of European countries. Of note, many of these and other studies used both DEA and SFA, and some also did comparisons to other approaches. As mentioned in the introduction, there are thousands of studies on this topic and these are just a few we mentioned, while many more examples can be found in Berger and Humphrey (1997), Fethi and Pasiouras (2010) and the recent Liu et al. (2013).

It is also critical to note that while a myriad of existing works are identified as 'DEA in banking' studies, many of them chose a particular variant of the many different variants within the DEA approach, besides also varying on the type of efficiency measures, as those described in the previous section, or their alternatives. It is, therefore, useful to understand at least the major variants within the DEA approach. Consequently, the goal of the rest of this section is to give a concise overview of different DEA models for efficiency and productivity analysis in general and for banking in particular. This overview is indeed brief, as covering it more extensively would take a book by itself, and these types of books already exist,[4] and here we give a primer that we hope will help learning from more detailed and much lengthier sources.

4.1 The Basic DEA Model

In their seminal work, Charnes et al. (1978) formulated a fractional programming problem for measuring efficiency of a decision making unit (DMU) with an allocation (x^j, y^j), using data on n observations in a sample, denoted as $\mathscr{S}_n = \{(x^k, y^k)\}_{k=1}^n$, for similar DMUs that can be considered as relevant peers for each other. This formulation (in our notation) was

[4]E.g., see Färe et al. (1994), Ray (2004) and most recent textbook-style treatment in Sickles and Zelenyuk (2019).

given by

$$\widehat{ITE}_{CCR}(x^j, y^j|\mathscr{S}_n) = \max_{\substack{v_1,...,v_N;\\ u_1,...,u_M}} \frac{\sum_{m=1}^{M} u_m y_m^j}{\sum_{l=1}^{N} v_l x_l^j} \tag{13}$$

$$s.t.$$

$$\frac{\sum_{m=1}^{M} u_m y_m^k}{\sum_{l=1}^{N} v_l x_l^k} \leq 1, k = 1, ..., n,$$
$$u_m \geq 0, m = 1, ..., M,$$
$$v_l \geq 0, l = 1, ..., N,$$

where $u' = (u_1, ..., u_M)$ and $v' = (v_1, ..., v_N)$ are optimization variables (also called 'multipliers').

To explain the objective function in (13) is the ratio of a weighted average of outputs to a weighted average of inputs for the observation (x^j, y^j). I.e., intuitively, it is a productivity index for the observation (x^j, y^j): the ratio of its aggregate output to its aggregate input, where the weights are obtained in the optimization problem (13), u and v, respectively. Moreover, it can also be seen as the so-called 'return to dollar' performance measure, i.e., the total revenue of this observation (x^j, y^j) divided by the total cost associated with (x^j, y^j), where the output and input prices are, again, obtained from the optimization problem (13), u and v, respectively. Furthermore, note the way these prices are obtained: the model searches for the best prices for the observation (x^j, y^j), in terms of the 'return to dollar' performance measure, subject to the constraints that this measure (with the same prices) is within $(0, 1]$ for all the observations in the sample. In this sense, the optimal values of u and v can be intuitively understood as (normalized) shadow prices that show (x^j, y^j) in the best possible light for the considered sample.

After transforming (13) to a linear programming (LP) problem, Charnes et al. (1978), obtained its dual formulation, which was the generalization (to multi-output case) AAM proposed by Farrell (1957), given by

$$\begin{aligned}
\widehat{ITE}(x^j, y^j|\mathscr{S}_n) &\equiv \min_{\theta, z^1,...,z^n} \{\theta \\
s.t. & \\
\sum_{k=1}^{n} z^k y_m^k &\geq y_m^j, m = 1, ..., M, \\
\sum_{k=1}^{n} z^k x_l^k &\leq \theta x_l^j, l = 1, ..., N, \\
\theta &\geq 0, z^k \geq 0, k = 1, ..., n\}
\end{aligned} \tag{14}$$

which is usually called the *'envelopment form of DEA'* assuming CRS, additivity and free disposability (see Sickles and Zelenyuk (2019) on how to derive the DEA formulation from these assumptions) and is more common in economics literature. Meanwhile, the formulation (13) is more common in the management science/operations research literature and is usually referred to as the 'multiplier form of DEA' (also under CRS, additivity and free disposability), or simply 'CCR model' to honor its authors.

Both formulations compute or estimate (from the sample) the input oriented Farrell technical efficiency measure because, as can be seen from (14), it minimizes inputs equiproportionately, while keeping outputs and the estimated technology fixed.

Similarly, the DEA-*estimator of the Farrell output oriented technical efficiency* for an observation (x^j, y^j), assuming CRS, free disposability of all outputs and all inputs and additivity, is given by

$$
\begin{aligned}
\widehat{OTE}(x^j, y^j | \mathscr{S}_n) &\equiv \max_{\lambda, z^1, ..., z^n} \{\lambda \\
& \quad s.t. \\
\sum_{k=1}^{n} z^k y_m^k &\geq \lambda y_m^j, m = 1, ..., M, \\
\sum_{k=1}^{n} z^k x_l^k &\leq x_l^j, l = 1, ..., N, \\
\lambda &\geq 0, z^k \geq 0, k = 1, ..., n\}
\end{aligned}
\tag{15}
$$

It is worth noting here that $\widehat{ITE}(x, y|\mathscr{S}_n) = 1/\widehat{OTE}(x, y|\mathscr{S}_n)$ for any (x, y) when they are obtained from (14) and (15), due to CRS, which is convenient and removes some ambiguity pertinent to the choice of one of these orientations.

Moreover, the estimates of the Shephard distance functions (Shephard (1953, 1970)), can be obtained by taking the reciprocals of the estimated Farrell efficiency measures.

If one is not willing to keep either inputs or outputs fixed and rather prefers expanding outputs and contracting inputs, simultaneously, then more general measures can be estimated (Sickles and Zelenyuk (2019, Chapter 8). For example, the directional distance function can also be estimated with DEA. E.g., when assuming CRS, free disposability of all outputs and all inputs and additivity, we can use:

$$
\begin{aligned}
\widehat{DDF}(x^j, y^j | d_x, d_y | \mathscr{S}_n) &\equiv \max_{\beta, z^1, ..., z^n} \{\beta, \\
s.t. & \\
\sum_{k=1}^{n} z^k y_m^k &\geq y_m^j + \beta d_{y_m},\ m = 1, ..., M, \\
\sum_{k=1}^{n} z^k x_l^k &\leq x_l^j - \beta d_{x_l},\ l = 1, ..., N, \\
& \quad z^k \geq 0,\ k = 1, ..., n\}.
\end{aligned}
\tag{16}
$$

Similarly, DEA is applicable to modeling the cost, revenue and profit functions and related efficiency measures. In particular, in such frameworks researchers estimate the overall (in)efficiency (e.g., cost, revenue or profit) and then decompose it into technical and allocative (in)efficiency parts. Sickles and Zelenyuk (2019, Chapter 3 and 8) elaborate on this in more details.

4.2 Other Variations of DEA

Many other types of DEA models were suggested in the literature, most of which are modifications of the basic models presented above. The modifications add various constraints to either the multiplier form of DEA, or the envelopment form.

4.2.1 Returns to Scale and Convexity

The first modifications tried to relax assumptions about CRS and convexity. The most popular among them is the DEA under the assumption of variable returns to scale (VRS) and, a bit less so, under the non-increasing returns to scale, while still assuming free disposability of all inputs and all outputs (and sub-additivity). These models impose additional constraints, $\sum_{k=1}^{n} z^k = 1$ or $\sum_{k=1}^{n} z^k \leq 1$, respectively, to the DEA-CRS formulation (14) or (15) or (16), depending on the chosen orientation or direction and, in general, these may yield different (and potentially very different) estimates. The multiplicative form of the DEA-VRS model is usually referred to as the BCC model due to Banker et al. (1984) (also see Afriat (1972); Färe et al. (1983)). For the context of banking these types of models were applied by Aly et al. (1990), Charnes et al. (1990) and Ferrier and Lovell (1990), English et al. (1993) and Wheelock and Wilson (2000), Fukuyama (1993); Fukuyama and Weber (2015), Camanho and Dyson (2005), Du et al. (2018), Casu et al. (2013), Curi et al. (2013, 2015), Simper et al. (2017), Casu et al. (2004), Koutsomanoli-Filippaki et al. (2009), Kenjegalieva et al. (2009a) to mention a few.

About the same time, the Free Disposal Hull (FDH) approach was developed and advocated by Deprins et al. (1984), where the idea is to relax the convexity (and, therefore, also additivity) and only keep the free disposability of all inputs and all outputs. While having its own name, FDH is a special case of DEA, where the envelopment form of DEA ((14) or (15) or (16), depending on the chosen orientation or direction) is amended into the DEA-VRS, but where the constraints "$z^k \geq 0, k = 1, ..., n$" are replaced with "$z^k \in \{0, 1\}, k = 1, ..., n$". As a result, it is a hybrid of LP and integer programming (IP) problems.

In the context of banking this type of approach was applied by Bauer and Hancock (1993) and Resti (1997). FDH typically has much lower discriminative power (and a slower rate of statistical convergence) and was used substantially less than DEA, though its popularity has increased more recently because of related work on conditional frontiers (e.g., see Daraio and Simar (2007)).

4.2.2 Modeling with Undesirable Outputs or with Congesting Inputs

All the models above assumed free disposability of inputs and outputs. This assumption is violated when there is a congestion phenomenon for inputs. E.g., when there are too many bank employees per square meter in an office, adding one more employee while keeping other inputs and technology fixed may even lead to lowering of the total output, implying that inputs are not freely disposable. Similarly, the free disposability of outputs is violated when there is an undesirable (or bad) output produced as a byproduct, along with a good output. In the context of banking, the nonperforming loan (NPL) is often considered as a typical example of a bad output. If the NPLs are known, some researchers simply subtract their total value from the total loans. An alternative way to handle them is to apply the DEA models that allow for weak disposability of outputs.

The roots for this type of modeling can be found in Shephard (1974), and more formalized in Färe and Svensson (1980); Färe and Grosskopf (1983); Grosskopf (1986); Tyteca (1996); Chung et al. (1997) and applied to various industries, with more recent developments in Seiford and Zhu (2002), Färe and Grosskopf (2003, 2004, 2009), Pham and Zelenyuk (2018, 2019) to mention a few.[5]

4.2.3 Other Streams of DEA

A number of streams of DEA modeling considered how to account for the network nature of production, both in static and dynamic contexts, starting from the seminal works of Färe and Grosskopf (1996); Färe et al. (1996) and many papers since then (e.g., see Kao (2014) for an excellent review).

Another stream of literature on DEA modeling focused on weight restrictions for the multiplier form of DEA: see Dyson and Thanassoulis (1988); Charnes et al. (1990); Thompson et al. (1990) and more recently Podinovski and Bouzdine-Chameeva (2013), and reviews from Allen et al. (1997); Podinovski (2015)).

Yet another promising stream of literature developed symbioses of DEA with game theory, and can be found in Hao et al. (2000); Nakabayashi and Tone (2006); Liang et al. (2008); Lozano (2012) and references therein.

Finally, a very important stream of DEA literature is about its statistical aspects—the wave largely developed by the seminal contributions by Léopold Simar and his co-authors, which we discuss in the next section.

4.3 Statistical Analysis of DEA and FDH

4.3.1 Convergence and Limiting distributions

The earliest proof of consistency of the basic DEA estimator, with a single-output for specification with output orientation was sketched by Banker (1993), who noticed that it is also a maximum likelihood estimator. Korostelev et al. (1995) advanced these results by proving more rigorous convergence of DEA and FDH estimators and also deriving their speeds of convergence (as orders of dimension of the production model), and proving their optimality properties. Kneip et al. (1998) generalized the theorems about convergence for the multi-input-multi-output case, Gijbels et al. (1999) discovered the limiting distribution of the DEA estimator (for the one-input-output case) and it was then generalized by Park et al. (2000), Jeong and Simar (2006), Kneip et al. (2008), Park et al. (2010).

4.3.2 Analysis of Distributions and Averages of DEA and FDH Estimates

A sub-stream of this literature has focused on analyzing the estimates of the DEA (or FDH) efficiency scores. Indeed, once such estimates are obtained it is useful, for example, to look at the estimated densities (at least histograms) of such scores, overall, or for some groups

[5] Also see Dakpo et al. (2017) for a good review of the research about undesirable outputs and congesting inputs.

within the sample and, in particular, test for the equality of distributions, as was explored in Simar and Zelenyuk (2006). For the context of banking, this type of approach was considered by Kenjegalieva et al. (2009b), Simper et al. (2017) and Du et al. (2018), to mention a few.

One can also analyze the averages of the efficiency scores, including weighted averages that account for an economic weight of each DMU in the aggregate, as were explored by Simar and Zelenyuk (2007) and more rigorously in Simar and Zelenyuk (2018). The latter work developed several new central limit theorems for aggregate efficiency, building upon the recent breakthrough due to Kneip et al. (2015). At about the same time, Kneip et al. (2016); Daraio et al. (2017) also applied the foundation from Kneip et al. (2015) to develop various statistical tests. Some further finite-sample improvements to these approaches (via improving the variance estimator) were developed by Simar and Zelenyuk (2020).

4.3.3 Regression Analysis with DEA and FDH Estimates

A very popular approach in DEA literature is the so-called 'two-stage DEA'. In the first stage the efficiency scores are estimated and, in the second stage, they are regressed on various explanatory factors that are believed to explain the variation in the efficiency scores. Early models applied OLS and then, to account for the 'boundedness' of the dependent variable, Tobit regression was often deployed. Currently, the state of the art here (albeit with its own caveats) is the method of Simar and Wilson (2007), who pointed out that (under certain conditions) truncated regression is more appropriate and the inference can be improved with the help of bootstrap.[6]

For the context of banking, these types of models were considered by Curi et al. (2013, 2015) and Du et al. (2018), to mention a few.

For example, a simple model that tries to explain (in terms of statistical dependencies) the distances between particular observations and the (estimate of the) *unconditional* frontier Ψ, can be stated as

$$\text{Efficiency}_i = f(Z_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{17}$$

where Z_i is a d-variate row-vector of explanatory variables (also called 'environmental' variables) for observation i, which are expected to be associated with the (in)efficiency score of this observation, which we denoted as Efficiency_i, via some functional form f, and ϵ_i is a statistical noise.

The type of the score, Efficiency_i, will determine the nature of the truncation. E.g., if output (input) oriented Farrell-type efficiency is applied, then both sides of (17) are bounded from below (above) by 1, implying that the distribution of ϵ_i is also bounded from below (above). Typically, one also makes parametric assumptions on f (e.g., linear in parameters) and the distribution of ϵ_i (e.g., truncated normal) and proceeds with parametric estimation using MLE.[7]

Because the true efficiency scores are unobserved and replaced by their estimates, which are inherently biased, the usual inference may be inaccurate, therefore Simar and Wilson (2007) proposed two alternative bootstrap algorithms that can (under certain conditions) mitigate these issues to some extent and improve the inference. If a researcher deals with

[6] Also see Simar and Wilson (2011).

[7] A nonparametric approach via local likelihood is also available, although with some degree of 'curse of dimensionality' and greater computational costs (e.g., see Park et al. (2008)).

panel data then the model 17 can be modified to exploit the richness of panel data, e.g., by estimating annual frontiers or accounting for a fixed or random effect. (E.g., see Du et al. (2018) and Sickles and Zelenyuk (2019, Chapter 10)).

Here it is important to emphasize that the analysis and the interpretation of this two-stage approach should be for the (in)efficiency scores with respect to the unconditional frontier, Ψ, that does not depend on the covariates Z_i by construction. At the same time, one should recognize that each DMU may have their own 'conditional' frontiers relative to which it might be even perfectly efficient. Recalling the analogy with an athletics competition: each athlete may have their own capacity (or 'technology') to perform at a given moment and the interest is to measure that performance not relative to their own capacity but relative to all other competitors. Alternatively, approaches like those in Daouia and Simar (2007); Bădin et al. (2012); Park et al. (2015); Simar et al. (2017) and related literature can be used if one is interested in measuring efficiency relative to conditional frontiers.

Concluding this section it is also worth clarifying that while the models of production, cost, revenue and profit functions are well-defined economic concepts where the researcher must know which variables to include, the selection of the 'environmental' variables is somewhat subjective, often varies across studies and therefore needs careful justification (see section 2.3 for the relevant discussion). Moreover, it is worth remembering that whether it is the conditional frontier approaches that are used or the truncated regression with respect to the unconditional frontier approaches, they are designed to estimate the statistical dependencies in the data under the assumed model and may or may not reveal causal relationships of the reality. Indeed, potential issues of endogeneity, reverse causality and selection bias may be relevant here (as for many other statistical analyses). Hence, integrating the existing methods to address these issues (e.g., such as those we briefly discuss in section 6) and their methods is a promising area for current and future research.

4.3.4 Noisy Data

A caveat of DEA and FDH is not explicitly handling noise in the data and hence a sensitivity to the so-called 'super-efficient' outliers. Various approaches were recommended to deal with this issue in the literature, from standard outlier-detection techniques to more sophisticated approaches tailored specifically to the DEA context. For example, Simar (2007); Simar and Zelenyuk (2011) proposed a formulation of Stochastic DEA (and Stochastic FDH). This approach goes in two stages:

Step 1. Filter the noise from the data with a non-parametric stochastic frontier analysis method (discussed in the next section), and then

Step 2. Apply DEA (or FDH) for the filtered data from Step 1.

All in all, DEA proved to be a very useful technique enabling a meaningful investigation of the bank efficiency in dimensions of types of inefficiencies, inefficiencies by bank divisions and input versus output orientation.

5 The Stochastic Frontier Analysis for Banking

SFA also gained significant popularity in research in general and on banking performance in particular. In a sense, it was inspired by DEA literature: it was developed as its competitor, its complement and also as its symbiotic partner. Like DEA, the SFA literature is vast and growing. It became a standard approach in the toolbox for the performance analysis of banks.

Among the first studies that applied SFA to banks was Ferrier and Lovell (1990), which we have already mentioned among the first DEA studies on banks, as they applied and compared both approaches in their basic, yet novel (at that time) forms. Being published in top econometrics journals, this work set the tone for many other future papers on SFA for banking and related fields. Among the first followers were Bauer and Hancock (1993), Berger (1993), Mester (1993), Akhavein et al. (1997), Berger and DeYoung (1997), Berger et al. (1997), Berger and Mester (1997), Berger and Hannan (1998), Hughes and Mester (1998), Adams et al. (1999), to mention a few among those that became classic for this literature.

More recent works include Kumbhakar et al. (2001), Casu et al. (2004), Kumbhakar and Tsionas (2005), Bos and Schmiedel (2007), Delis and Tsionas (2009), Casu et al. (2013), Malikov et al. (2016), to mention a few.[8] Again, these are just a few examples from a thousands of studies that are impossible to mention all here, though an interested reader can find many more examples in reviews by Berger and Humphrey (1997), Fethi and Pasiouras (2010).

Similarly to DEA, while thousands of existing works are identified as 'SFA in banking' studies, many of them deployed particular variations out of the many within the SFA approach, as well as variations on the type of efficiency they looked at. Thus, the goal of this section is to provide a concise review of the different methods within the SFA approach. Again, the review is aimed to be brief, because covering it in more detail would take several chapters, or a whole book, which already exists.[9]

5.1 The Basic SFA Model

The first stochastic frontier models were developed at about the same time by Aigner et al. (1977) (hereafter ALS) and Meeusen and van den Broeck (1977). The idea was basically to add one more term to the standard regression model—the inefficiency term, u, unobserved by researchers and random, with a one-sided asymmetric distribution for it (e.g., half normal or exponential). This term is in addition to the usual error term with a symmetric (e.g., normal) distribution. It was first suggested in the single output cross-sectional framework, where the output is modeled as

$$y_i = \psi(x_i) \exp\{\varepsilon_i\}, \quad i = 1, ..., n \tag{18}$$

where

$$\psi(x) = \max\{y : (x, y) \in \Psi\}$$

is the deterministic frontier, or production function (for any fixed x), to be estimated and ε_i is a composed-error, defined as a convolution of the usual and unobserved statistical error,

[8] Note that some of these and other studies used particular versions of both DEA and SFA.

[9] A classic book here is Kumbhakar and Lovell (2000), though many more developments came after it and the most recent summary to date can be found, for example in Sickles et al. (2020), as a very short practical review, and more thorough review in Kumbhakar et al. (2020a,b), while a comprehensive textbook style treatment can be found in Chapter 11 through 16 of Sickles and Zelenyuk (2019), which also connects to the economic theory foundations and DEA described in prior chapters.

or 'white noise', v_i, and an unobserved inefficiency term $u_i \geq 0$ that forces the DMU i to produce below the frontier output, i.e.,

$$\varepsilon_i = v_i - u_i. \tag{19}$$

Thus, the efficiency of firm i can be measured by

$$\text{Efficiency}_i = \exp\{-u_i\} \equiv \frac{y_i}{\psi(x_i)\exp\{v_i\}}, \quad i = 1, ..., n \tag{20}$$

and often approximated (around 1) as $1 - u_i$, and hence u_i approximates the inefficiency of DMU i.

The model is usually estimated in natural logs, i.e.,

$$\log y_i = \log \psi(x_i) + v_i - u_i, \quad i = 1, ..., n \tag{21}$$

most commonly with some parametric assumptions on $\log \psi(\cdot)$, e.g., linear in logs (i.e., $\psi(\cdot)$ is assumed to be Cobb-Douglas) or Translog, or any other suitable functional form, although more recent developments proposed various semi and non-parametric generalizations (e.g., see Sickles and Zelenyuk (2019, Chapter 16) and Parmeter and Zelenyuk (2019) for more detailed comparisons).

While many distributions can be used for u and v, the most common is the original specification from Aigner et al. (1977), where

$$v_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$$

and

$$u_i \overset{iid}{\sim} |\mathcal{N}(0, \sigma_u^2)|$$

and also assuming that v_i and u_i are independent from each other and from x_i. In turn, this implies that ε_i is an iid random variable, independent from x_i, with a density given by

$$f_\varepsilon(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right], \quad -\infty \leq \varepsilon \leq +\infty \tag{22}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively, while $\lambda = \sigma_u/\sigma_v$ is sometimes referred to as 'signal-to-noise' ratio and $\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$ and then the (approximate) average inefficiency is given by

$$E(u_i) = \sqrt{\frac{2}{\pi}}\sigma_u = \left(\frac{2}{\pi(1 + \lambda^2)}\right)^{1/2} \sigma\lambda. \tag{23}$$

The model can then be estimated with a standard maximum likelihood estimator (MLE) or even with a method of moments estimator (MME) due to Olson et al. (1980), which will generate estimates of β, λ and σ, which in turn can be used to obtain estimates of σ_v and σ_u. Plugging the estimate of σ_u or (σ and λ) into (23) will give an estimate of the (approximate) average inefficiency. The model can also be re-parametrized to obtain estimates of $E(\exp\{-u_i\})$, although estimation of (23) appears to be more common in practice.

The estimates of the individual inefficiencies are usually proxied by the so-called JLMS-estimator, due to Jondrow et al. (1982), which estimate $E(u_i|\varepsilon_i)$ for specific distributions of u_i and v_i. Specifically, when u_i and v_i are half-normal and normal, respectively, then Jondrow et al. (1982) showed that

$$E(u_i|\varepsilon_i) = \frac{\sigma_v \sigma_u}{\sigma} \left[-\frac{\varepsilon_i \lambda}{\sigma} + \frac{\phi\left(\varepsilon_i \lambda / \sigma\right)}{1 - \Phi(\varepsilon_i \lambda / \sigma)} \right], \tag{24}$$

where the unobserved quantities $\sigma_v, \sigma_u, \varepsilon_i, \lambda, \sigma$ can be replaced with their MLE or MM estimates to obtain the estimates of $E(u_i|\varepsilon_i)$ for each observation $i = 1, ..., n$. Again, it is also possible to re-parametrize the model to obtain $E(\exp\{-u_i\}|\varepsilon_i)$. [10]

Finally, here by convention we outlined SFA for the case of production frontier estimation and a similar logic (with some modifications) applies to the estimation cost frontier, revenue frontier and profit frontier approaches. In these contexts, researchers try to estimate the overall (in)efficiency (e.g., cost, revenue or profit efficiency) and then decompose it into technical and allocative (in)efficiency, which may require estimation of a system of equations.[11]

5.2 Panel Stochastic Production Frontiers

Often researchers have panel data, where each observation i is observed over several periods, $t = 1, ..., T$. One can still employ the ALS approach described above by pooling the panel and treating it as a cross section (i.e., hence the name 'pooled SFA'). While this is a good start, it is usually beneficial to exploit the richness of the panel data by using SFA approaches tailored specifically for panel data, and we briefly describe some of them below. More details can be found in Sickles and Zelenyuk (2019, Chapter 11-15) and its brief version in Sickles et al. (2020).

5.2.1 Basic Panel Data SFA

The first attempts at the SFA for panel data go back to Pitt and Lee (1981) and Schmidt and Sickles (1984). In a nutshell, they modeled the frontier as

$$y_{it} = \alpha + x_{it}\beta + v_{it} - u_i, \;\; i = 1, ..., n; \; t = 1, ..., T, \tag{25}$$

where x_{it} is the row vector of N inputs used by firm i in period t, and then re-parametrized it, by letting $\alpha_i = \alpha - u_i$, as

$$y_{it} = \alpha_i + x_{it}\beta + v_{it}, \;\; i = 1, ..., n; \; t = 1, ..., T. \tag{26}$$

Observing (26), one can see it looks exactly like a standard panel data regression model, which can be estimated (and tested) using the pooled OLS approach, the fixed effects (FE) approach, or the random effects (RE) approach.

[10] E.g., see Sickles and Zelenyuk (2019, Chapter 11).

[11] E.g., for detailed discussion on this see the classic work of Kumbhakar (1997) and some more recent developments in Kumbhakar (1997); Kumbhakar and Tsionas (2005); Mamatzakis et al. (2015); Malikov et al. (2016) as well as a textbook-style discussion in Sickles and Zelenyuk (2019).

In the case of the FE estimation approach, one can obtain the estimates of the individual fixed effects, α_i, call them $\widehat{\alpha}_i$, and then find their maximum value in the sample, $\widehat{\alpha} \equiv \max_i\{\widehat{\alpha}_i\}$, to serve as the benchmark and then define estimates of inefficiency scores as

$$\widehat{u}_i = \widehat{\alpha} - \widehat{\alpha}_i \geq 0, \quad i = 1, ..., n$$

i.e., this means that the most efficient DMU in the sample is assigned 0% inefficiency by construction.

All approaches have their own caveats and the main one for this approach is that the estimates of (in)efficiency are assumed to be time-invariant. Moreover, with the nature of the FE approach in general, all the time-invariant heterogeneity pertinent to the individuals is absorbed by the estimates of FEs, which all feed into the estimates of efficiency scores.

5.2.2 Other Panel Data SFA Approaches

Various modifications of SFA for panel data were later suggested in the literature, mainly to account more for the unobserved heterogeneity at various levels, including allowing for time-varying inefficiency. For example, Cornwell et al. (1990), started with the model

$$y_{it} = Z_i\alpha + x_{it}\beta + w_{it}\delta_i + v_{it} \quad i = 1, ..., n;\ t = 1, ..., T, \tag{27}$$

where Z_i and w_{it} are row vectors of explanatory factors affecting y_{it} via corresponding column vectors of parameters α and δ_i, besides the affects from the inputs x_{it} via the column vector β.

They then re-parametrized model, by letting $\delta_i = \delta_0 + u_i$ where $\delta_0 = E[\delta_i]$, as following:

$$y_{it} = Z_i\alpha + x_{it}\beta + w_{it}\delta_0 + \varepsilon_{it},$$

where ε_{it} is a composed error generalization of (19), given by

$$\varepsilon_{it} = v_{it} + w_{it}u_i \tag{28}$$

and it is assumed that u_i is *i.i.d.* random variable with a finite positive-definite covariance matrix Ω while the statistical noise term, v_{it}, is *i.i.d.* with zero mean and constant variance σ_v^2 and is uncorrelated with Z_i, x_{it}, and u_i. Furthermore, various time-dependency structures can then be assumed, e.g., by setting $w_{it} = (1, t, t^2)$. They also discussed various strategies of estimating this model, which were adaptations of Fixed-effects estimator, Random-effects estimator and Hausman-Taylor estimator, depending on the assumed correlation structure between the errors and regressors.

Other variations of panel data SFA were also proposed in many other interesting works, most notably in Kumbhakar (1990), Lee (1991) and Lee and Schmidt (1993), Battese and Coelli (1995). More recent elaborations on the panel data SFA can be found in Greene (2005a,b), Ahn et al. (2007) and one of the most general approaches by Kneip et al. (2012) (which is analogous to seminar works of Bai and Ng (2002) and Bai (2009)). Also see Colombi et al. (2011), Colombi et al. (2014), Tsionas and Kumbhakar (2014) and a special issue in Journal of Econometrics edited by Kumbhakar and Schmidt (2016) for many interesting discussions.

5.3 Explaining Inefficiency

Special attention has been paid to many SFA models (especially those mentioned in previous subsection) and how to explain the inefficiency with some of the explanatory variables. The work of Battese and Coelli (1995) seems to be very popular on this, in practical work, probably due to its simplicity and availability of software for it from the early days. The idea of this approach is to model the frontier as

$$y_{it} = X_{it}\beta + v_{it} - u_{it} \quad i = 1, ..., n;\ t = 1, ..., T, \tag{29}$$

with additional assumption that u_{it} is not purely random but has some regularities or deterministic part, e.g.,

$$u_{it} = Z_{it}\delta + w_{it}, \tag{30}$$

where Z_{it} and is a row vector of explanatory factors (also called 'environmental' variables) believed to explain u_{it} via the corresponding column vector of parameters δ, up to some noise w_{it}. Because of the requirement that $u_{it} \geq 0$, truncation restriction must be imposed on w_{it}, namely $w_{it} \geq -Z_{it}\delta$ and some distribution assumed to disentangle it from other error terms. They specifically considered

$$w_{it} \sim \mathcal{N}(0, \sigma_w^2), \qquad w_{it} \geq -Z_{it}\delta \ i = 1, ..., n;\ t = 1, ..., T \tag{31}$$

which (as they noted) also equivalent to stating that

$$u_{it} \sim \mathcal{N}(Z_{it}\delta, \sigma_u^2),\ s.t.\ u_{it} \geq 0,\ i = 1, ..., n;\ t = 1, ..., T. \tag{32}$$

Since the model is fully parametrized, one likelihood function can be constructed after substituting (34) into (29) and can be optimized to get the ML estimates of $\beta, \sigma_w, \sigma_v$ and δ and their standard errors (from the Fisher information matrix). The estimate of δ will indicate the association of u_{it} on z_{it} in the sense of (34) and significance tests can be done in the usual way in econometrics (e.g., using t-tests or LR-tests).

Here it is worth noting that some researchers also tried to analyze SFA efficiency scores in two stages, first estimating JLMS efficiency scores in ALS framework and then regress them onto explanatory variables. While this may seem natural, it is incoherent in the ALS framework: if there is a belief that inefficiency scores $u_i, i = 1, ..., n$ are not iid and they depend on some characteristics then the assumptions of ALS are violated and instead one should incorporate those beliefs into the model and the corresponding likelihood structure, e.g., as in Battese and Coelli (1995), etc.[12]

While very intuitive and relatively simple, it is important to realize that the approach of Battese and Coelli (1995) often suffers from computational problems and numerical identification issues in particular. One reason for the problems is because $Z_{it}\delta$, while believed to be part of the inefficiency, is influencing y_{it} in essentially the same way as $x_{it}\beta$ (except for the impact through truncation on w_{it}, which can be minimal). A way to mitigate this problem is to model the inefficiency through the second (rather than the first) moment of u, that is, via σ_u, e.g.,

[12] More discussion on this can be found in Wang and Schmidt (2002). Also see Kim and Schmidt (2008) for various statistical testing issues in this context have.

$$u_{it} \sim \mathcal{N}(0, \sigma_u^2(Z_{it})),\ s.t.\ u_{it} \geq 0,\ i = 1, ..., n;\ t = 1, ..., T$$

with some structure assumed for $\sigma_u^2(Z_{it})$ that makes sure it is non-negative.[13] For example, one can assume

$$\log(\sigma_u^2(Z_{it})) = Z_{it}\delta, \quad i = 1, ..., n;\ t = 1, ..., T \tag{33}$$

and then estimate this relationship in the MLE framework and obtain fitted values of $\sigma_u^2(Z_{it})$ for particular Z_{it} of interest (e.g., observed points), and then, in this (half-normal) case, one can estimate

$$E(u_{it}|z_{it}) = \sqrt{\frac{2}{\pi}}\sigma_u(Z_{it}), \quad i = 1, ..., n;\ t = 1, ..., T \tag{34}$$

which (under certain assumptions) will give estimates of the individual inefficiency scores for any particular Z_{it}.

Concluding this section it is also worth reiterating the caveats, which are similar to those we already mentioned for the analogous DEA context. Namely, while the models of production, cost, revenue and profit functions are well-defined economic concepts where the researcher must know which variables to include, the selection of the 'environmental' variables is more subjective, may vary across studies and thus demands careful justifications (e.g., see section 2.3 for some examples). Moreover, these and most of SFA approaches are also designed to estimate the statistical dependencies in the data under the assumed statistical model, which may or may not reveal the causal relationship of the reality. Therefore, adapting various methods for mitigating potential issues of endogeneity, reverse causality and selection bias may be relevant here (e.g., such as those we briefly discuss in section 6) to SFA methods and is a promising area for current and future research. Some of the fundamental works on this include Horrace et al. (2016), Kumbhakar and Tsionas (2016), Simar et al. (2016)[14] and more recently Amsler et al. (2017), to mention just a few.

5.4 Other Variations of SFA

All the SFA models above were parametric in the sense that parametric assumptions on the functional form for the frontier had to be made (in addition to parametric assumptions on the inefficiency and noise in some cases). More recent literature has tried to avoid such assumptions. Among the first works on this are the papers by Fan et al. (1996) and Kneip and Simar (1996), who suggested kernel based methods. The latter work was also considered in the panel data context, which was soon complemented by the research of Adams et al. (1999) and Park et al. (2003).

More recent elaborations came due to Park et al. (2007), which employed the local likelihood approach, and a similar approach by Martins-Filho and Yao (2015), and further generalizations by Park et al. (2015). These approaches estimated the frontier non-parametrically,

[13] This approach is a special case of the nonparametric approach developed by Park et al. (2015) in local-likelihood framework and Simar et al. (2017) in local-least squares framework.

[14] Also see other papers in the same Special Issue as these papers and the introduction from the Editors, Kumbhakar and Schmidt (2016).

yet required local parametric assumptions on the inefficiency and the statistical noise. A somewhat simpler to compute (and with less assumptions) approach was more recently proposed by Simar et al. (2017). The latter work also developed a novel way of estimating the marginal effects of covariates onto expected inefficiency without imposing any parametric assumptions, exploiting the advantages of the one-parameter distribution family that is natural to assume for the inefficiency term. For more discussions of these methods, see Parmeter and Zelenyuk (2019).

Finally, as in virtually any other field involving statistics, various Bayesian approaches for SFA have been developed. Among the classical works on this are due to Koop et al. (1994, 1995, 1997, 1999) and more recently by Griffin and Steel (2007), Tsionas and Papadakis (2010) and Liu et al. (2017), to mention a few. For the context of banking these types of models were considered by Kumbhakar and Tsionas (2005) and Malikov et al. (2016), to mention examples.

Overall, SFA proves to be an evolving technique enabling both aggregation of data in the focus of analysis and breakdown into identified specifications of cost and profit efficiencies, that could be enriched with the conditional variables, specified by researcher.

6 Other Econometric Approaches

Recent finance literature on banking performance dedicated significant attention to causal inference. The roots of the methods of causal evaluation take their origins from the seminal works of Splawa-Neyman et al. (1990) and Fisher (1925) on statistical tests of treatment effects in randomized experiments. Methods of causal inference currently thrive in microeconometrics, in particular, in labour and health economics (Bertrand et al. (2004)), and other areas (Imbens and Wooldridge (2009)).

Recent applications of these methods in banking include Laeven and Levine (2009), Aiyar et al. (2014), Duchin and Sosyura (2014), Nanda and Nicholas (2014), Schepens (2016), Anginer et al. (2018), Buchak et al. (2018), Neuhann and Saidi (2018), Zelenyuk et al. (2019), Zelenyuk et al. (2020), to mention a few.

The literature on casual inference is fairly large by itself and it seems infeasible to cover it even briefly within the few remaining pages of this chapter. Instead, we recommend the reader to get familiar with these methods through already written excellent books, e.g., by Manski (2009), Angrist and Pischke (2009), Imbens and Rubin (2015), to mention a few that also provide many references. Our goal here, therefore, is to discuss a few recent examples of how the causal analysis can be applied in the context of performance of banks, and to provide references for further details.

6.1 Importance of Causal Inference in Analysis of Banking Performance

Causal inference has received significant attention in the recent banking literature on the topics of the performance effects that change after the passage of regulatory laws, or rules. Most interesting econometric questions would consider the cause and the resulting effect triggered by the new rule. Popular empirical questions here would be of a nature as to whether the new capital requirements introduced by the regulatory rule affect the riskiness of bank loans. The research question is often about the drivers (e.g., see Berger and Bouwman (2013), Acharya and Mora (2015)) that could have a causal effect on the performance of banks. For example, Bonner and Eijffinger (2015) find the causal effect of liquidity requirements on bank interest margins. In another study, Hamadi et al. (2016) find that capital increase due to Basel rules increases market valuation of loan loss provisions. Both outputs of interest margins and loan loss provisions are in attention of studies of bank performance, because they contribute to the major banking function of lending.

To address these and similar policy type questions with econometric methods a theoretical model is needed. A model is basically a framework for inter-relationships between parameters. If these parameters can be identified, than the model is estimable (Cameron and Trivedi (2010)). In the presence of economically and statistically significant results, conclusions about the policy effects would be valuable for the literature and for the policy makers. To derive relevant conclusions the most interesting model set up needs to identify the causal and 'effect' parameters and control for any observed confounding variables. If the confounding factors are observed (and controlled for) the conclusions could be derived about causality. In the event, when confounders are unobserved, causal effects could be derived with the instrumental variables techniques (Angrist and Pischke (2009)).

The key three questions in the modeling are about the event (or treatment), the cause-effect relationship and the confounding factors. In application to banking, causal inference is often driven by an exogenous (beyond control of the bank) event. Rarely, this event is entirely random (global crisis, pandemics, etc.), yet more often it is a policy intervention (regulation of bankers pay gaps, etc.). In an empirical framework, any of these events can create an experimental setting, where a relevant question is about the effects that this event brings on the parameters of interest, in particular, effects on the lending of banks, etc. The literature examples include the effects of bail-outs (Duchin and Sosyura (2014)) and capital enhancement (Berger and Bouwman (2013)) policies on bank performance, where the bank performance is presented with loans and market share, respectively.

Investigation of the effects of the policy enhancements with the research methods involves identification of the banks that were impacted by the policy (Zelenyuk et al. (2019)), the counterfactual (not impacted), and identification of the impact-related parameters. The impact of an event could vary among banks because of a different assignment to the policy rules and the differential response of banks to the rules. For example, capital enhancing policies could affect only banks of a certain size in a certain state. The banks of the same size in a similar state could form a control group. And the affected parameter could be bank lending.

Of critical importance for an empirical design is whether the banks are self-selecting by the results of an event, whether they are assigned to comply with the rule (assigned to 'treatment'), and whether the assignment is random. If a regulator has selected banks to a mandatory capital adequacy increase, an assignment to treatment is beyond the control of the bank and the effect of the policy can be evaluated with the difference-in-differences methods (DD). If the change in law is random, causal methods of regression discontinuity design (RDD) could be applied. Alternatively, if banks are self-selecting to treatment, for example, banks are deciding on the provision of information about their capital adequacy and they inform the market voluntarily, then difference-in-differences methods may be inappropriate for causal inference. For these types of problems, the instrumental variable (IV) approach can be more relevant, yielding consistent estimates, if the underlying regularity conditions hold. Fuzzy versus sharp RDD design are applied in similar situations to IV methods.

We now embark on the task to briefly present the essence of each of these methods in their application to studying bank performance.

6.2 The Role of DD in Causal Inference on Bank Performance

DD is one of the working models in modern econometric analysis and its detailed description can be found in many textbooks, e.g., in Cameron and Trivedi (2010), Greene (2011), Wooldridge (2013) and Imbens and Rubin (2015), and papers, including by Bertrand et al. (2004), Imbens and Wooldridge (2008), that we follow here. In its idea, DD method is a comparison between the outcomes post and pre an event versus the difference to the outcomes of similar banks that are unaffected by the event. The emphasis is on a 'causal parameter', impacting the resulting outcome. The rest of the variables are considered as controls. More specifically, with the assumption that a policy changing parameter δ measures the impact on the treated sample and X is a conditional variable of the treatment, an outcome variable Y can be formally described via

$$Y_{it} = X_{it}\beta + \delta D_{it} + \varepsilon_{it} \tag{35}$$

In a simple regression, comparing treated and untreated samples, β_2 measures the difference in average outcomes pre and post intervention for a treated group.

$$Y_{it} = \beta_1 + \beta_2 D_{it} + \varepsilon_{it}, \tag{36}$$

where the average outcome conditional on $D_{it} = 0$, of those who did not receive treatment, is compared to an average outcome conditional on $D_{it} = 1$, of treated.

In a banking example, Duchin and Sosyura (2014) focused on the effect of government aid on risk taking by banks. Specifically, the authors studied the difference between the 'before' versus 'after' the government assistance approval for the qualifying banks in the difference to a number of control banks: unapproved, those that did not apply for aid and eligible banks. The range of variables of interest affected with the government capital program included retail mortgages, corporate loans and securities. Interestingly, the authors find no significant effect of the government aid programs on the loan volume, yet they find that the loans originated by the approved banks are riskier than the loans of any of the control groups of banks. The findings are based on a linear probability regression, where an outcome variable is 1/0 for approved/unapproved loans. The authors found that the approved for aid banks increased their new credit origination by 5.4 percentage points for riskier mortgages, as well as for syndicated loans and riskier securities. To deal with possible selection bias and to demonstrate that the treatment effect is due to a government program rather than due to a selection of approved banks, the authors apply propensity score matching and the instrumental variables techniques.

In a further example, the effect is of the capital adequacy increase on lending. Here, lending is an outcome variable and capital adequacy is distinguished between the treated –

required to increase the capital adequacy of banks and untreated (control) banks that do not need to increase their capital adequacy. The difference between the capital adequacy of the treated versus the capital adequacy of control banks after the policy intervention would contribute to the determination of the average treatment effect resulting from a policy event. For a more complete analysis of the policy effect the researcher would need to evaluate the difference post versus pre policy intervention. The sample is comprised of four groups: the treatment group before the policy change, the treatment group (the one required to increase capital adequacy) after the policy change, the control group before the policy change and the control group (not required to increase capital adequacy) after the policy change.

Formally, let Y_{it} be the outcome variable of interest, lending for bank i in period t, then the model can be specified, similarly to Zelenyuk et al. (2019), as follows

$$Y_{it} = \alpha_0 + \beta_1 C_{it} + \beta_2 Post_t + \beta_3 C_{it} Post_t + X_{it}\gamma + \varepsilon_{it} \tag{37}$$

where

$Post_t$ is a variable which is equal to zero before the capital increasing reform and one after the reform.

C_{it} is a variable which equals one if the bank is in the treatment group and is required to increase its capital after the reform milestone and zero otherwise.

X_{it} is a row vector of control variables and γ is the corresponding column vector of parameters. A usual guide for selecting bank-level controls is the CAMELS system (see Table 1 above), where variables are considered to explain bank performance remarkably well (e.g., see Eichler et al. (2011) for the related discussion).

β_3 is the DD coefficient that captures the difference between the lending outcomes for the treatment group and the control group post reform. The estimate of β_3 provides the average treatment effect because it measures the effect of the policy of capital increase on the average outcome of lending.

ε_{it} is an error (i.e., unexplained) term, assumed to satisfy certain regularity conditions (e.g., see Cameron and Trivedi (2010), Wooldridge (2013), Imbens and Rubin (2015) for the formal details and the related discussions of the limitations, as well as Cameron and Trivedi (2010) for implementation in Stata).

Overall, and as with any approach, the causalities identified with the average treatment effects have certain caveats. In particular, for any effect to be causal, the governmental policies need to be unexpectedly assigned to the banks, while the treatment and control groups need to be similar by their key characteristics before the policy changing event. Some of these challenges in estimation could be dealt with application of RDD methods, others can be approached with the instrumental variables techniques. Both of these methods are discussed in sub-sections 6.3 and 6.4, respectively.

6.3 The Role of RDD in Causal Inference on Bank Performance

In cases when randomized assignment to treatment is impossible, yet there is an effect of a policy change that is 'as good as random', an alternative to DD strategy is RDD. The foundations of this method go back to at least Thistlethwaite and Campbell (1960) and are further described and elaborated by Angrist and Pischke (2009), with the famous application by Hahn et al. (2001) and a practical guide by Imbens and Lemieux (2007), Cameron and Trivedi (2010) and Greene (2011), which we follow here.

In its essence, the RDD approach allocates observations to a treatment and a control group based on the value of an assignment variable, exogenous to an experiment. Examples of an assignment variable include poverty rate and its effect on educational programs and class size on examination performance Greene (2011). Continuing the example in sub-section 6.2. of the capital increase, suppose the treatment is assigned based on the systemic importance of a bank. In particular, in macroeconomics literature a bank is often considered to be systemically important if its failure might be a trigger to a financial crisis (Adrian and Brunnermeier (2016)), which in turn implies a threshold on certain variables, e.g., assets of a bank.

When the assignment to treatment is activated by the crossing of a fixed threshold, i.e., the size of a bank, the effects of the treatment can be evaluated with the RDD. Discontinuity stems from different values of the outcome variable depending on whether the bank is above or below the treatment threshold, or a certain cut-off point. It is assumed that the banks closest to the cutoff are similar by their key characteristics. In our example, the lending growth could be different for the increasing capital banks versus the non-increasing capital banks when the assignment to treatment is determined based on crossing the cutoff. The idea here is that the banks cannot manipulate the assignment variable, that is the banks cannot change their size, or systemic importance, over a very short period of time, then the treatment occurs with certainty and the policy effects of the capital increase on lending can be evaluated.

This assignment to treatment feature of the regression discontinuity design allows to match the banks relatively well. Importantly, the matching is based on the cutoff characteristics rather than on any other complex method of propensity score estimator, or the 'nearest neighbor' estimator[15]. RDD presents an alternative to the matching experiment (Thistlethwaite and Campbell (1960)), when there is an equivalency in treatment and control groups prior to a policy event. Alternatively, RDD allocates observations to the treatment and control groups conditional on the value of a predictor to the left and to the right sides of a cutoff (Imbens and Lemieux (2007)). The caveat of RDD is that the predictor might be associated with the outcomes, yet, if the association is functionally continuous (only an outcome variable is discontinuous), any distributional discontinuity of the outcome variable, conditional on the predicting variable at the cut-off point is subject to a causality interpretation, provided that the regularity conditions on the data generating process hold (e.g., see Imbens and Lemieux (2007) for further details).

Critically, any estimated treatment effect with RDD has "local" properties. It is an assumption that the banks are well matched around the cutoff, the further an observation

[15] For more details on propensity score matching and nearest neighbour estimator, see Rubin and Thomas (1996), Rubin (2006), while for practical implementation, see Cameron and Trivedi (2010).

is from the cutoff the more different the banks are expected to be. How well the banks are matched is testable. The outcome variable would be continuous without the policy intervention and discontinuity arises as a result of the policy event. Therefore, it is expected that the effect is causal and dominates the effect of the other variables, except for an identified relationship between the outcome (Y) and the forcing (X) variable.

To identify causality, an analysis needs to find the discontinuity (Imbens and Lemieux (2007)). An interesting example is the recent study about liquidity enhancing policies by Bonner and Eijffinger (2015), who found that the interest rates distribution is discontinuous around the liquidity buffer cutoff.

After the discontinuity is identified, a test is needed to estimate whether the difference of the outcome variable is significant. In the capital increase example, if X_i is a dummy variable for whether a bank needs to increase capital with a new policy and Z_i is a variable representing the systemic importance of the bank based on size, then:

$$\begin{aligned} X_i &= 1, \quad \text{if } Z_i \geq Z_0 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

where Z_0 is the systemic important threshold.

Here, the RDD implementation is an estimation of two regressions of outcome variable Y_i (lending) on X_i in samples to the left and to the right of the size cutoff. A comparison of the intercepts from these two regressions estimates a change in the lending variable Y_i given treatment X_i. For further flexibility of implementation, the model can be formulated in general terms as:

$$Y_i = \alpha(Z_i) + X_i\beta + \varepsilon_i \tag{38}$$

where $\alpha_i = \alpha(Z_i) + \varepsilon_i$ is a function of Z_i and other unobservables ε_i, with the assumptions that $E[\alpha_i|Z_i = Z]$ is continuous in Z at Z_0, while the density of Z_i is positive in the neighborhood of Z_0, along with other regularity conditions that allow identification and consistent estimation and valid inference for this model (see Hahn et al. (2001) for details).

To conclude, in the application of the RDD, researchers need to be aware of its caveats. In particular, the estimated treatment effects are likely to have 'local' features due to an analysis of observations that are placed close to a cutoff.

6.4 Performance Analysis in the Presence of Selection Bias

Selection bias presents a significant challenge in research, aiming to conduct causal inference. An impossibility of observing the outcome after the policy change in a difference to the outcome for the same banks should the policy change not occur, creates a fundamental problem of causal inference, described by Holland (1986). Selection bias is essentially the difference between the outcome of the treatment banks should the policy change not occur and the outcome of the control banks. Randomization in the 'ideal' experiment resolves the selection bias questions. However, in the presence of selection bias, an estimate of an average treatment effect may be inconsistent. For an extensive review we recommend Holland (1986), Heckman (1990), Heckman et al. (1997), Imbens and Wooldridge (2009), Angrist and

Pischke (2009), Cameron and Trivedi (2010). In performance of banks literature, Allen et al. (1990), Cantor and Packer (1997), Maskara and Mullineaux (2011) deal with selection bias by changing the sample, inclusion of the possible determinants of the resulting variable, and the IV type estimation, respectively.

To summarize, two types of selection bias are most common (Imbens and Wooldridge (2009), Cameron and Trivedi (2010)), these are selection on observable and selection on unobservable variables. In selection bias on observables, the treatment variable is correlated with the error in the outcome equation due to an omitted observable variable that determines both the treatment and outcome variable. In selection bias on unobservables, the correlation between the treatment variable and the error in the outcome equation is due to an omitted unobservable variable that partly determines both treatment and outcome variables.

To deal with the selection bias problem, in the first event a common approach is to include all the observable variables (potentially correlated with an error term in the outcome equation) into an outcome equation and then estimate this equation with least squares. In the event of selection on unobservables, a usual approach is to remedy the selection bias problem in a two-stage procedure, parametric or semiparametric estimation.

Recently Zelenyuk et al. (2020) adapted a two-stage approach of Heckman (1976) and Maddala (1983) – to deal with the selection bias in a study of the impact of voluntary disclosure of capital adequacy on bank lending in the U.S. In a two-stage procedure the model includes an outcome variable, treatment assignment and instrumental variables.[16] To be more precise, in their model, when banks self-select to voluntary disclosure, treatment of disclosure on lending can be endogenous and is assumed to follow the process

$$Y_i = X_i\beta + aD_i + \varepsilon_i \tag{39}$$

where Y_i is the lending adjusted for past performance (essentially a growth of lending) for bank i, X_i is a row vector of control variables and β is the corresponding column vector of parameters, while D_i is a treatment participation decision variable that equals 1 or 0. Moreover, the treatment participation indicator depends on the instrumental variable Z, via the following specification

$$D_i^* = \gamma_0 + \gamma_1 Z_i + v_i \tag{40}$$

where D_i^* is a latent variable with an observable counterpart D_i defined as 1/0.

Importantly, the variable Z is in the equation of D (40), but not in the equation of Y in (39). Leveraging on the linear nature of the model, assuming $cov[Z, v] = 0, cov[\varepsilon, Z] = 0, cov[X, \varepsilon] = 0$, and $cov[D, Z] \neq 0$, and other assumptions on random errors v_i and the data generating process in general, it can be shown that a consistently estimates the average treatment effect (e.g., see Cameron and Trivedi (2010) and references therein).

An indicator about treatment participation depends on the instrumental variable. Zelenyuk et al. (2020) selected an instrument based on the theory of voluntary disclosure (Diamond and Dybvig (1983)) and corporate governance (Shleifer and Vishny (1997)). In particular, they chose stock-based managerial compensation as an instrument because of the role of the stock-related compensation in dealing with the agency problems. Then, the two-stage

[16] For related discussions, also see Cameron and Trivedi (2010).

algorithm was applied to the model for an estimation of probability of voluntary disclosure and the effect of this disclosure on the lending. Here an estimate of a is an IV type estimate that has local properties (Gippel et al. (2015)). The local average treatment effect depends on Z being applied in the treatment evaluation and on an instrument. With this approach, Zelenyuk et al. (2020) found significant evidence of a positive effect of a voluntary bank capital adequacy disclosure on lending.

Finally, it is important to note that a limitation of the IV application to an inference involving selection bias is that treatment and control groups may not be representative of the whole population, therefore, the results of these types of estimations may not be robustly supporting big policy questions, yet they could still be informative of more local changes, such as policy thresholds, etc.

7 Concluding Remarks

In this chapter, we overview major approaches of performance analysis in banking. These approaches include envelopment-type estimators (DEA and FDH, in many variations), stochastic frontier estimators (parametric, semiparametric and nonparametric) and other non-efficiency type econometric methods. We provide a relatively brief overview of each stream and cite many (although, of course, not all) works we believe would be useful for the interested readers to find more details about them.

It might be also worth emphasizing that the inclusion of the last part, briefly describing methods of DD, RDD, IV, is a key difference of our chapter relative to most (if not all) other works on performance analysis in banking and, perhaps, overall that we are aware of to date. The reason for including this interesting area of research in our review, however, is not just to distinguish our work from others. The goal is to bring attention of both audiences, which have so far largely developed on their own. This is because, indeed, we strongly believe a lot of interesting research should result from the synthesis and the synergies of these approaches, both for new theoretical developments and for future interesting practical applications.

Finally, there are also other areas of performance analysis that we have not covered here. One of them that so far has been largely under-explored and we believe has a fruitful future is the adaptation of methods in machine learning and artificial intelligence (including the contexts of Big Data and Social Networks) for performance analysis in general and for performance of banking in particular.[17] We hope and encourage this to be addressed with future research endeavors.

[17] E.g., see Mullainathan and Spiess (2017), Athey and Imbens (2019) and Zelenyuk (2020) and Ya Chen and Zelenyuk (2020) for some recent examples of such studies for performance analysis.

References

Acharya, V. V. and Mora, N. (2015). A crisis of banks as liquidity providers. *The Journal of Finance*, 70(1):1–43. 1, 2.3, 6.1

Adams, R. M., Berger, A. N., and Sickles, R. C. (1999). Semiparametric approaches to stochastic panel frontiers with applications in the banking industry. *Journal of Business and Economic Statistics*, 17(3):349–358. 5, 5.4

Adrian, T. and Brunnermeier, M. K. (2016). CoVar. *American Economic Review*, 106(7):1705–41. 2.3, 6.3

Afriat, S. N. (1972). Efficiency estimation of production functions. *International Economic Review*, 13(3):568–598. 4.2.1

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2007). Stochastic frontier models with multiple time-varying individual effects. *Journal of Productivity Analysis*, 27(1):1–12. 5.2.2

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37. 1.1, 5.1, 5.1

Aiyar, S., Calomiris, C. W., Hooley, J., Korniyenko, Y., and Wieladek, T. (2014). The international transmission of bank capital requirements: Evidence from the uk. *Journal of Financial Economics*, 113(3):368 – 382. 6

Akhavein, J. D., Berger, A. N., and Humphrey, D. B. (1997). The effects of megamergers on efficiency and prices: evidence from a bank profit function. *Finance and Economics Discussion Series 1997-9, Board of Governors of the Federal Reserve System*. 5

Allen, D. S., Lamy, R. E., and Thomson, G. R. (1990). The shelf registration of debt and self selection bias. *The Journal of Finance*, 45(1):275–287. 6.4

Allen, R., Athanassopoulos, A., Dyson, R. G., and Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research*, 73:13–34. 4.2.3

Almanidis, P., Karakaplan, M. U., and Kutlu, L. (2019). A dynamic stochastic frontier model with threshold effects: U.S. bank size and efficiency. *Journal of Productivity Analysis*, 52(2):69 – 84. 2.2, 2.3

Altunbas, Y., Carbo, S., Gardener, E. P., and Molyneux, P. (2007). Examining the relationships between capital, risk and efficiency in European banking. *European Financial Management*, 13(1):49–70. 2.3

Aly, H. Y., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, scale, and allocative efficiencies in U.S. banking: An empirical investigation. *The Review of Economics and Statistics*, 72(2):211–218. 2.2, 2.3, 4, 4.2.1

Amsler, C., Prokhorov, A., and Schmidt, P. (2017). Endogenous environmental variables in stochastic frontier models. *Journal of Econometrics*, 199(2):131–140. 5.3

Anginer, D., Demirguc-Kunt, A., Huizinga, H., and Ma, K. (2018). Corporate governance of banks and financial stability. *Journal of Financial Economics*, 130. 1.2, 1, 6

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, Princeton University Press. 6, 6.1, 6.3, 6.4

APRA (2013). Domestic Systemically Important Banks in Australia. *Australian Prudential Regulation Authority Information Paper*. 2.3

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 60(1):47–57. 1.2

Ashenfelter, O. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4):648–60. 1.2

Assaf, A. G., Berger, A. N., Roman, R. A., and Tsionas, M. G. (2019). Does efficiency help banks survive and thrive during financial crises? *Journal of Banking and Finance*, 106:445 – 470. 1, 2.3

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685725. 17

Bădin, L., Daraio, C., and Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3):818–833. 4.3.3

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279. 5.2.2

Bai, J., Carvalho, D., and Phillips, G. M. (2018). The impact of bank credit on labor reallocation and aggregate industry productivity. *The Journal of Finance*, 73(6):2787–2836. 1

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221. 5.2.2

Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science*, 39(10):1265–1273. 4.3.1

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092. 4.2.1

Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2):325–332. 5.2.2, 5.3, 5.3

Bauer, P. W. and Hancock, D. (1993). The efficiency of the Federal Reserve in providing check processing services. *Journal of Banking and Finance*, 17(2-3):287–311. 4.2.1, 5

BCBS (2019). An examination of initial experience with the global systemically important bank framework. *Banking Commission on Banking Supervision Bank for International Settlements Working Paper 34*. 2.3

Berg, T. (2018). Got Rejected? Real Effects of Not Getting a Loan. *The Review of Financial Studies*, 31(12):4912–4957. 1.2

Berger, A., DeYoung, R., Genay, H., and Udell, G. F. (2000). Globalization of financial institutions: Evidence from cross-border banking performance. *Brookings-Wharton Papers on Financial Services*, 28(7):23–120. 2.3

Berger, A. N. (1993). “distribution-free” estimates of efficiency in the us banking industry and tests of the standard distributional assumptions. *Journal of productivity Analysis*, 4(3):261–292. 5

Berger, A. N. and Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109(1):146 – 176. 2.3, 6.1

Berger, A. N. and Bouwman, C. H. S. (2009). Bank Liquidity Creation. *The Review of Financial Studies*, 22(9):3779–3837. 1

Berger, A. N. and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance*, 21(6):849 – 870. 2.3, 5

Berger, A. N. and Hannan, T. H. (1998). The efficiency cost of market power in the banking industry: A test of the “quiet life” and related hypotheses. *Review of Economics and Statistics*, 80(3):454–465. 5

Berger, A. N., Hasan, I., and Zhou, M. (2010). The effects of focus versus diversification on bank performance: Evidence from Chinese banks. *Journal of Banking and Finance*, 34(7):1417 – 1435. 2.3

Berger, A. N. and Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, 98(2):175 – 212. 2.2, 4, 5

Berger, A. N., Leusner, J. H., and Mingo, J. J. (1997). The efficiency of bank branches. *Journal of Monetary Economics*, 40(1):141 – 162. 2.2, 5

Berger, A. N. and Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance*, 21(7):895 – 947. 2.3, 5

Berger, A. N. and Roman, R. A. (2017). Did saving Wall street really save MainÂ street? the real effects of TARP on local economic conditions. *Journal of Financial and Quantitative Analysis*, 52(5):1827–1867. 1

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates?*. *The Quarterly Journal of Economics*, 119(1):249–275. 6

Bikker, J. and Bos, J. (2008). Bank performance: A theoretical and empirical framework for the analysis of profitability, competition, and efficiency. *Bank Performance: A Theoretical and Empirical Framework for the Analysis of Profitability, Competition and Efficiency*. 1, 2.1, 2.3

BIS (2015). Making supervisory stress tests more macroprudential: Considering liquidity and solvency interactions and systemic risk. *Bank for International Settlements Working Paper 29, https://www.bis.org*. 2.1

BIS (2020a). The Basel Framework. *Bank for International Settlements, https://www.bis.org*. 1, 2.3

BIS (2020b). History of the Basel Committee. *Bank for International Settlements, https://www.bis.org/bcbs/history.htm*. 1

Bonner, C. and Eijffinger, S. C. W. (2015). The Impact of Liquidity Regulation on Bank Intermediation. *Review of Finance*, 20(5):1945–1979. 1.2, 2.4, 6.1, 6.3

Bos, J. and Schmiedel, H. (2007). Is there a single frontier in a single European banking market? *Journal of Banking and Finance*, 31(7):2081–2102. 5

Boubakri, N., El Ghoul, S., Guedhami, O., and Hossain, M. (2020). Post-privatization state ownership and bank risk-taking: Cross-country evidence. *Journal of Corporate Finance*, 64:101625. 2.4

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453 – 483. 1.2, 6

Camanho, A. S. and Dyson, R. G. (2005). Cost efficiency measurement with price uncertainty: A dea application to bank branch assessments. *European Journal of Operational Research*, 161(2):432–446. 2.2, 4, 4.2.1

Cameron, C. and Trivedi, P. (2010). *Microeconometrics Using Stata: Revised Edition*. Stata Press. 2.4, 6.1, 6.2, 6.2, 6.3, 15, 6.4, 6.4, 16

Cantor, R. and Packer, F. (1997). Differences of opinion and selection bias in the credit rating industry. *Journal of Banking Finance*, 21(10):1395 – 1417. 6.4

Casu, B., Ferrari, A., and Zhao, T. (2013). Regulatory reform and productivity change in Indian banking. *The Review of Economics and Statistics*, 95(3):1066–1077. 2.2, 4, 4.2.1, 5

Casu, B., Girardone, C., and Molyneux, P. (2004). Productivity change in European banking: A comparison of parametric and non-parametric approaches. *Journal of Banking and Finance*, 28(10):2521 – 2540. 1, 4, 4.2.1, 5

Casu, B., Girardone, C., and Molyneux, P. (2006). *Introduction to Banking*. Prentice Hall Financial Times. 2.1

Chambers, R., Chung, Y., and Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, 98(2):351–364. 3

Chambers, R. G., Chung, Y., and Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory*, 70(2):407–419. 3

Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444. 1.1, 4.1, 4.1

Charnes, A., Cooper, W. W., Huang, Z. M., and Sun, D. B. (1990). Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of econometrics*, 46(1-2):73–91. 4, 4.2.1, 4.2.3

Chung, Y. H., Färe, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51(3):229–240. 4.2.2

Colombi, R., Kumbhakar, S. C., Martini, G., and Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis*, 42(2):123–136. 5.2.2

Colombi, R., Martini, G., and Vittadini, G. (2011). A stochastic frontier model with short-run and long-run inefficiency random effects. University of Bergamo Department of Economics and Technology Management Working Paper No. 012011. 5.2.2

Cornwell, C., Schmidt, P., and Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46(1):185–200. 5.2.2

Curi, C., Guarda, P., Lozano-Vivas, A., and Zelenyuk, V. (2013). Is foreign-bank efficiency in financial centers driven by home or host country characteristics? *Journal of Productivity Analysis*, 40(3):367–385. 4, 4.2.1, 4.3.3

Curi, C., Lozano-Vivas, A., and Zelenyuk, V. (2015). Foreign bank diversification and efficiency prior to and during the financial crisis: Does one business model fit all? *Journal of Banking and Finance*, 61(S1):S22–S35. 4, 4.2.1, 4.3.3

Dakpo, K. H., Jeanneaux, P., and Latruffe, L. (2017). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research*, 250(2):347–359. 5

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375 – 400. 4.3.3

Daraio, C. and Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis*, 28(1):13–32. 4.2.1

Daraio, C., Simar, L., and Wilson, P. W. (2017). Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production. *The Econometrics Journal*. 4.3.2

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3):273–292. 1.1

Delis, M. D. and Tsionas, E. G. (2009). The joint estimation of bank-level market power and efficiency. *Journal of Banking and Finance*, 33(10):1842 – 1850. 5

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labour efficiency in post offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurement*, pages 243–267. Amsterdam, NL: Springer. 4.2.1

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419. 1, 6.4

Du, K., Worthington, A. C., and Zelenyuk, V. (2018). Data envelopment analysis, truncated regression and double-bootstrap for panel data with application to Chinese banking. *European Journal of Operational Research*, 265(2):748 – 764. 4, 4.2.1, 4.3.2, 4.3.3, 4.3.3

Duchin, R. and Sosyura, D. (2014). Safer ratios, riskier portfolios: Banks' response to government aid. *Journal of Financial Economics*, 113(1):1–28. 1.2, 2.1, 1, 2.4, 6, 6.1, 6.2

Dyson, R. G. and Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 39(6):563–576. 4.2.3

EBA (2020). Global Systemically Important Institutions. *European Banking Authority*. 2.3

Eichler, S., Karmann, A., and Maltritz, D. (2011). The term structure of banking crisis risk in the united states: A market data based compound option approach. *Journal of Banking and Finance*, 35(4):876 – 885. Crete Conference 2010: The Future of Universal Banking. 2.4, 6.2

Eling, M. and Luhnen, M. (2010). Efficiency in the international insurance industry: A cross-country comparison. *Journal of Banking and Finance*, 34(7):1497 – 1509. 2.2

English, M., Grosskopf, S., Hayes, K., and Yaisawarng, S. (1993). Output allocative and technical efficiency of banks. *Journal of Banking and Finance*, 17(2-3):349–366. 1.1, 4, 4.2.1

Fan, Y., Li, Q., and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics*, 14(4):460–468. 5.4

Färe, R. and Grosskopf, S. (1983). Measuring congestion in production. *Zeitschrift für Nationalökonomie / Journal of Economics*, 43(3):257–271. 4.2.2

Färe, R. and Grosskopf, S. (1996). *Intertemporal Production Frontiers: With Dynamic DEA*. Norwell, MA: Kluwer Academic Publishers. 4.2.3

Färe, R. and Grosskopf, S. (2003). Nonparametric productivity analysis with undesirable outputs: Comment. *American Journal of Agricultural Economics*, 85(4):1070–1074. 4.2.2

Färe, R. and Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation: Comment. *European Journal of Operational Research*, 157(1):242–245. 4.2.2

Färe, R. and Grosskopf, S. (2009). A comment on weak disposability in nonparametric production analysis. *American Journal of Agricultural Economics*, 91(2):535–538. 4.2.2

Färe, R., Grosskopf, S., and Logan, J. (1983). The relative efficiency of Illinois electric utilities. *Resources and Energy*, 5(4):349–367. 4.2.1

Färe, R., Grosskopf, S., and Lovell, C. A. K. (1994). *Production Frontiers*. New York, NY: Cambridge University Press. 4

Färe, R., Grosskopf, S., and Roos, P. (1996). On two definitions of productivity. *Economics Letters*, 53(3):269–274. 4.2.3

Färe, R., He, X., Li, S. K., and Zelenyuk, V. (2019). A unifying framework for Farrell efficiency measurement. *Operations Research*, 67(1):183–197. 2.1, 3.2

Färe, R. and Svensson, L. (1980). Congestion of production factors. *Econometrica: Journal of the Econometric Society*, 48(7):1745–1753. 4.2.2

Farrell, M. J. (1957). The measurement of productive efficiency. *Jornal of the Royal Statistical Society. Series A (General)*, 120(3):253–290. 1.1, 4.1

Federal Reserve (2020). BHC supervision manual. *Board of Governors of the Federal Reserve System, www.federalreserve.gov*. 2.1, 2.3

Ferrier, G. and Lovell, C. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics*, 46(1-2):229–245. 1.1, 4, 4.2.1, 5

Fethi, M. D. and 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. 2.2, 4, 5

Fisher, R. (1925). Statistical methods for research workers. *Oliver and Boyd, Edinburgh*. 1.2, 6

Fukuyama, H. (1993). Technical and scale efficiency of Japanese commerical banks: a non-parametric approach. *Applied Economics*, 25(8):1101–1112. 1.1, 4, 4.2.1

Fukuyama, H. and Weber, W. (2015). Measuring Japanese bank performance: a dynamic network DEA approach. *Journal of Productivity Analysis*, 44(3):249–264. 2.3, 4, 4.2.1

Georgescu-Roegen, N. (1951). The aggregate linear production function and its applications to von Neumann's economic model. In Koopmans, T., editor, *Activity Analysis of Production and Allocation*. New York, NY: Wiley. 2.1

Gijbels, I., Mammen, E., Park, B. U., and Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, 94(445):220–228. 4.3.1

Gippel, J., Smith, T., and Zhu, Y. (2015). Endogeneity in accounting and finance research: Natural experiments as a state-of-the-art solution. *Abacus*, 51(2):143–168. 6.4

Goodhart, C. (2011). The Basel Committee on Banking Supervision: A History of the Early Years 1974-1997. *Cambridge University Press*. 1

Greene, W. H. (2005a). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, 23(1):7–32. 5.2.2

Greene, W. H. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303. 5.2.2

Greene, W. H. (2011). *Econometric Analysis*. Prentice Hall, 7th edition. 6.2, 6.3

Griffin, J. E. and Steel, M. F. (2007). Bayesian stochastic frontier analysis using winbugs. *Journal of Productivity Analysis*, 27(3):163–176. 5.4

Grosskopf, S. (1986). The role of the reference technology in measuring productive efficiency. *The Economic Journal*, 96(382):499–513. 4.2.2

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209. 1.2, 6.3, 6.3

Hamadi, M., Heinen, A., Linder, S., and Porumb, V.-A. (2016). Does Basel II affect the market valuation of discretionary loan loss provisions? *Journal of Banking and Finance*, 70:177 – 192. 6.1

Hao, G., Wei, Q. L., and Yan, H. (2000). A game theoretical model of DEA efficiency. *The Journal of the Operational Research Society*, 51(11):1319–1329. 4.2.3

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5:475–492. 6.4

Heckman, J. (1990). Varieties of selection bias. *American Economic Review*, 80(2):313–18. 6.4

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654. 6.4

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960. 6.4

Horrace, W. C., Liu, X., and Patacchini, E. (2016). Endogenous network production functions with selectivity. *Journal of Econometrics*, 190(2):222–232. 5.3

Hughes, J. and Mester, L. (1998). Bank capitalization and cost: Evidence of scale economies in risk management and signaling. *The Review of Economics and Statistics*, 80(2):314–325. 1.1, 5

Humphrey, D. (2020). Distance functions, bank output, and productivity. *Journal of Productivity Analysis*, pages 13–26. 2.2

Imbens, G. and Lemieux, T. (2007). Regression Discontinuity Designs: A guide to practice. NBER Working Papers 13039, National Bureau of Economic Research, Inc. 1.2, 6.3

Imbens, G. and Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press. 6, 6.2, 6.2

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86. 6, 6.4

Isik, I. and Hassan, M. (2002). Technical, scale and allocative efficiencies of Turkish banking industry. *Journal of Banking and Finance*, 26(4):719 – 766. 2.2, 2.3

Jeong, S.-O. and Simar, L. (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis*, 97(10):2141–2161. 4.3.1

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2–3):233–238. 5.1

Kao, C. (2014). Network data envelopment analysis: A review. *European Journal of Operational Research*, 239(1):1–16. 4.2.3

Kenjegalieva, K., Simper, R., Weyman-Jones, T., and Zelenyuk, V. (2009a). Comparative analysis of banking production frameworks in Eastern European financial markets. *European Journal of Operational Research*, 198(1):326 – 340. 2.2, 4, 4.2.1

Kenjegalieva, K., Simper, R., Weyman-Jones, T., and Zelenyuk, V. (2009b). Comparative analysis of banking production frameworks in Eastern European financial markets. *European Journal of Operational Research*, 198(1):326–340. 4.3.2

Kim, M. and Schmidt, P. (2008). Valid tests of whether technical inefficiency depends on firm characteristics. *Journal of Econometrics*, 144(2):409–427. 12

Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14(6):783–793. 4.3.1

Kneip, A., Sickles, R. C., and Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, 28(3):590–628. 5.2.2

Kneip, A. and Simar, L. (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis*, 7(2–3):187–212. 5.4

Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24(6):1663–1697. 4.3.1

Kneip, A., Simar, L., and Wilson, P. W. (2015). When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory*, 31(2):394–422. 4.3.2

Kneip, A., Simar, L., and Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business & Economic Statistics*, 34(3):435–456. 4.3.2

Koop, G., Osiewalski, J., and Steel, M. F. (1994). Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function. *Journal of Business & Economic Statistics*, 12(3):339–346. 5.4

Koop, G., Osiewalski, J., and Steel, M. F. (1997). Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers. *Journal of Econometrics*, 76(1):77–105. 5.4

Koop, G., Osiewalski, J., and Steel, M. F. J. (1999). The components of output growth: A stochastic frontier analysis. *Oxford Bulletin of Economics and Statistics*, 61(4):455–487. 5.4

Koop, G., Steel, M. F., and Osiewalski, J. (1995). Posterior analysis of stochastic frontier models using Gibbs sampling. *Computational Statistics*, 10:353–373. 5.4

Koopmans, T. (1951). *Activity Analysis of Production and Allocation*. New York, NY: Wiley. 1.1

Korostelev, A., Simar, L., and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Annals of Statistics*, 23(2):476–489. 4.3.1

Koutsomanoli-Filippaki, A., Margaritis, D., and Staikouras, C. (2009). Efficiency and productivity growth in the banking industry of Central and Eastern Europe. *Journal of Banking and Finance*, 33(3):557 – 567. 2.3, 4, 4.2.1

Kumbhakar, S. and Schmidt, P. (2016). *Endogeneity Problems in Econometrics, Special Issue of the Journal of Econometrics*, volume 190. Amsterdam, Netherlands: Elsevier. 5.2.2, 14

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(1):201–211. 5.2.2

Kumbhakar, S. C. (1997). Modeling allocative inefficiency in a translog cost function and cost share equations: An exact relationship. *Journal of Econometrics*, 76(1):351–356. 11

Kumbhakar, S. C. and Lovell, C. A. K. (2000). Knox (2000) Stochastic Frontier Analysis. 9

Kumbhakar, S. C., Lozano-Vivas, A., Lovell, C. A. K., and Hasan, I. (2001). The effects of deregulation on the performance of financial institutions: The case of Spanish savings banks. *Journal of Money, Credit and Banking*, 33(1):101–120. 5

Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2020a). *Stochastic Frontier Analysis: Foundations and Advances I*, page forthcoming. Springer Singapore, Singapore. 9

Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2020b). *Stochastic Frontier Analysis: Foundations and Advances II*, page forthcoming. Springer Singapore, Singapore. 9

Kumbhakar, S. C. and Tsionas, E. G. (2005). Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. *Journal of Econometrics*, 126(2):355 – 384. 5, 11, 5.4

Kumbhakar, S. C. and Tsionas, E. G. (2016). The good, the bad and the technology: Endogeneity in environmental production models. *Journal of Econometrics*, 190(2):315–327. 5.3

Laeven, L. and Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93(2):259 – 275. 6

Leary, M. T. (2009). Bank loan supply, lender choice, and corporate capital structure. *The Journal of Finance*, 64(3):1143–1185. 2.3

Lee, Y. H. (1991). *Panel data models with multiplicative individual and time effects: Applications to compensation and frontier production functions*. PhD thesis, Michigan State University, East Lansing, MI, USA. 5.2.2

Lee, Y. H. and Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency: Techniques and Applications*, pages 237–255. New York, NY: Oxford University Press. 5.2.2

Leontief, W. (1925). Die Bilanz der Russischen Volkswirtschaft - Eine methodologische Untersuchung. *Weltwirtschaftliches Archiv*, 22(2):338–344. 1.1

Leverty, J. T. and Grace, M. F. (2010). The robustness of output measures in property-liability insurance efficiency studies. *Journal of Banking and Finance*, 34(7):1510 – 1524. 2.2

Liang, L., Wu, J., Cook, W. D., and Zhu, J. (2008). The DEA game cross-efficiency model and its Nash equilibrium. *Operations Research*, 56(5):1278–1288. 4.2.3

Liberman, A. (2016). The value of a good credit reputation: Evidence from credit card renegotiations. *Journal of Financial Economics*, 120(3):644 – 660. 1.2

Liu, J., Sickles, R. C., and Tsionas, E. G. (2013). Bayesian treatments to panel data models. Unpublished manuscript, Rice University, Houston, TX. 4

Liu, J., Sickles, R. C., and Tsionas, E. G. (2017). Bayesian treatments to panel data models with time-varying heterogeneity. *Econometrics*, 5(33):1–21. 5.4

Lozano, S. (2012). Information sharing in DEA: A cooperative game theory approach. *European Journal of Operational Research*, 222(3):558–565. 4.2.3

Lozano-Vivas, A., Pastor, J. T., and Pastor, J. M. (2002). An efficiency comparison of European banking systems operating under different environmental conditions. *Journal of Productivity Analysis*, 18(1):59–77. 2.3, 4

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press. 6.4

Malikov, E., Kumbhakar, S. C., and Tsionas, M. G. (2016). A cost system approach to the stochastic directional technology distance function with undesirable outputs: The case of US banks in 2001 - 2010. *Journal of Applied Econometrics*, 31(7):1407–1429. 5, 11, 5.4

Mamatzakis, E., Matousek, R., and Vu, A. N. (2019). What is the impact of problem loans on Japanese bank productivity growth? *Financial Markets, Institutions & Instruments*, 28(2):213–240. 2.2, 2.3

Mamatzakis, E., Tsionas, M. G., Kumbhakar, S. C., and Koutsomanoli-Filippaki, A. (2015). Does labour regulation affect technical and allocative efficiency? Evidence from the banking industry. *Journal of Banking and Finance*, 61:S84 – S98. 11

Manski, C. F. (2009). *Identification for Prediction and Decision*. Princeton, Princeton University Press. 6

Martinez-Peria, M. S. and Schmukler, S. (2001). Do depositors punish banks for bad behavior? Market discipline, deposit insurance, and banking crises. *The Journal of Finance*, 56(3):1029–1051. 1, 2.4

Martins-Filho, C. and Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews*, 34(4):413–451. 5.4

Maskara, P. K. and Mullineaux, D. J. (2011). Information asymmetry and self-selection bias in bank loan announcement studies. *Journal of Financial Economics*, 101(3):684 – 694. 6.4

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–44. 1.1, 5.1

Mester, L. J. (1993). Efficiency in the savings and loan industry. *Journal of Banking and Finance*, 17(2–3):267–286. 1, 5

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106. 17

Nakabayashi, K. and Tone, K. (2006). Egoist's dilemma: a DEA game. *Omega*, 34(2):135–148. 4.2.3

Nanda, R. and Nicholas, T. (2014). Did bank distress stifle innovation during the great depression? *Journal of Financial Economics*, 114(2):273 – 292. 6

Neuhann, D. and Saidi, F. (2018). Do universal banks finance riskier but more productive firms? *Journal of Financial Economics*, 128(1):66 – 85. 6

Olson, J. A., Schmidt, P., and Waldman, D. M. (1980). A Monte Carlo study of estimators of the stochastic frontier production function. *Journal of Econometrics*, 13(1):67–82. 5.1

Paradi, J. C. and Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega*, 41(1):61 – 79. Data Envelopment Analysis: The Research Frontier - This Special Issue is dedicated to the memory of William W. Cooper 1914-2012. 2.2

Park, B. U., Jeong, S.-O., and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Annals of Statistics*, 38(3):1320–1340. 4.3.1

Park, B. U., Sickles, R. C., and Simar, L. (2003). Semiparametric-efficient estimation of AR(1) panel data models. *Journal of Econometrics*, 117(2):279–309. 5.4

Park, B. U., Sickles, R. C., and Simar, L. (2007). Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics*, 136(1):281–301. 5.4

Park, B. U., Simar, L., and Weiner, C. (2000). The FDH estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16(6):855–877. 4.3.1

Park, B. U., Simar, L., and Zelenyuk, V. (2008). Local likelihood estimation of truncated regression and its partial derivatives: Theory and application. *Journal of Econometrics*, 146(1):185–198. 7

Park, B. U., Simar, L., and Zelenyuk, V. (2015). Categorical data in local maximum likelihood: Theory and applications to productivity analysis. *Journal of Productivity Analysis*, 43(2):199–214. 4.3.3, 5.4, 13

Parmeter, C. F. and Zelenyuk, V. (2019). Combining the virtues of stochastic frontier and data envelopment analysis. *Operations Research*, 67(6):1628–1658. 5.1, 5.4

Pham, M. D. and Zelenyuk, V. (2018). Slack-based directional distance function in the presence of bad outputs: Theory and application to Vietnamese banking. *Empirical Economics*, 54(1):153–187. 4.2.2

Pham, M. D. and Zelenyuk, V. (2019). Weak disposability in nonparametric production analysis: A new taxonomy of reference technology sets. *European Journal of Operational Research*, 274(1):186 – 198. 2.3, 4.2.2

Pitt, M. M. and Lee, L.-F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics*, 9(1):43–64. 5.2.1

Podinovski, V. V. (2015). DEA models with production trade-offs and weight restrictions. In Zhu, J., editor, *Data Envelopment Analysis: A Handbook of Models and Methods*, pages 105–144. New York, NY: Springer. 4.2.3

Podinovski, V. V. and Bouzdine-Chameeva, T. (2013). Weight restrictions and free production in data envelopment analysis. *Operations Research*, 61(2):426–437. 4.2.3

Ray, S. C. (2004). *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. New York, NY: Cambridge University Press. 4

Resti, A. (1997). Evaluating the cost-efficiency of the Italian banking system: What can be learned from the joint application of parametric and non-parametric techniques. *Journal of Banking and Finance*, 21(2):221 – 250. 4.2.1

Roncalli, T. (2020). *Handbook of Financial Risk Management*. New York: Chapman and Hall/CRC. 2.1, 2.1

Ross, T., Van de Venter, B., and Westerfield, J. (2017). *Essentials of Corporate Finance*. McGraw-Hill Education (Australia). 2.1

Rubin, D. (2006). *Matched Sampling for Causal Effects*. Cambridge University Press, Cambridge, UK. 15

Rubin, D. and Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*, 52(1):249–264. 15

Schepens, G. (2016). Taxes and bank capital structure. *Journal of Financial Economics*, 120(3):585 – 600. 6

Schmidt, P. and Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, 2(4):367–374. 5.2.1

Sealey Jr., C. W. and Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance*, 32(4):1251–1266. 1, 2.2

Seiford, L. M. and Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, 142(1):16–20. 4.2.2

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton, NJ: Princeton University Press. 1.1, 4.1

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton studies in mathematical economics. Princeton, NJ: Princeton University Press. 4.1

Shephard, R. W. (1974). Indirect production functions. In *Mathematical Systems in Economics*, volume 10. Hain, Meisenheim am Glan. 4.2.2

Sherman, H. D. and Gold, F. (1985). Bank branch operating efficiency: Evaluation with data envelopment analysis. *Journal of Banking and Finance*, 9(2):297–315. 1.1, 4

Shleifer, A. and Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, 52(2):737–783. 2.3, 6.4

Sickles, R. and Zelenyuk, V. (2019). *Measurement of Productivity and Efficiency: Theory and Practice*. New York, NY: Cambridge University Press. 3.1, 3.2, 3.3, 4, 4.1, 4.1, 4.1, 4.3.3, 9, 5.1, 5.2, 10, 11

Sickles, R. C., Song, W., and Zelenyuk, V. (2020). Chapter 8 - econometric analysis of productivity: Theory and implementation in R. In Vinod, H. D. and Rao, C., editors, *Financial, Macro and Micro Econometrics Using R*, volume 42 of *Handbook of Statistics*, pages 267 – 297. Elsevier. 9, 5.2

Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis*, 28(3):183–201. 4.3.4

Simar, L., Van Keilegom, I., and Zelenyuk, V. (2017). Nonparametric least squares methods for stochastic frontier models. *Journal of Productivity Analysis*, 47(3):189–204. 4.3.3, 13, 5.4

Simar, L., Vanhems, A., and Van Keilegom, I. (2016). Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. *Journal of Econometrics*, 190(2):360 – 373. Endogeneity Problems in Econometrics. 5.3

Simar, L. and Wilson, P. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64. 2.3, 3.1, 4.3.3, 4.3.3

Simar, L. and Wilson, P. W. (2011). Two-stage DEA: Caveat emptor. *Journal of Productivity Analysis*, 36(2):205–218. 3.1, 6

Simar, L. and Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews*, 25(4):497–522. 4.3.2

Simar, L. and Zelenyuk, V. (2007). Statistical inference for aggregates of Farrell-type efficiencies. *Journal of Applied Econometrics*, 22(7):1367–1394. 4.3.2

Simar, L. and Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, 36(1):1–20. 4.3.4

Simar, L. and Zelenyuk, V. (2018). Central limit theorems for aggregate efficiency. *Operations Research*, 166(1):139–149. 4.3.2

Simar, L. and Zelenyuk, V. (2020). Improving finite sample approximation by central limit theorems for estimates from Data Envelopment Analysis. *European Journal of Operational Research*, 284(3):1002–1015. 4.3.2

Simper, R., Hall, M. J. B., Liu, W., Zelenyuk, V., and Zhou, Z. (2017). How relevant is the choice of risk management control variable to non-parametric bank profit efficiency analysis? The case of South Korean banks. *Annals of Operations Research*, 250(1):105–127. 4, 4.2.1, 4.3.2

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4):465–472. 1.2, 6

Sturm, J.-E. and Williams, B. (2004). Foreign bank entry, deregulation and bank efficiency: Lessons from the Australian experience. *Journal of Banking and Finance*, 28(7):1775–1799. 2.3

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51:309–317. 1.2, 6.3

Thompson, R. G., Langemeier, L. N., Lee, C.-T., Lee, E., and Thrall, R. M. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, 46(1-2):93–108. 4.2.3

Trezevant, R. (1992). Debt financing and tax status: Tests of the substitution effect and the tax exhaustion hypothesis using firms' responses to the economic recovery tax act of 1981. *The Journal of Finance*, 47(4):1557–1568. 1.2

Tsionas, E. G. and Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics*, 29(1):110–132. 5.2.2

Tsionas, E. G. and Papadakis, E. N. (2010). A Bayesian approach to statistical inference in stochastic DEA. *Omega*, 38(5):309–314. 5.4

Tyteca, D. (1996). On the measurement of the environmental performance of firms—a literature review and a productive efficiency perspective. *Journal of Environmental Management*, 46(3):281–308. 4.2.2

Von Neumann, J. (1945). A model of general equilibrium. *Review of Economic Studies*, 13(1):1–9. 1.1

Wang, H.-J. and Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, 18(2):129–144. 3.1, 12

Wheelock, D. C. and Wilson, P. W. (2000). Why do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions. *The Review of Economics and Statistics*, 82(1):127–138. 4, 4.2.1

Wooldridge, J. (2013). *Introductory Econometrics: A Modern Approach*. Mason, OH: South-Western, 5th edition. 6.2, 6.2

Ya Chen, M. T. and Zelenyuk, V. (2020). Lasso+dea for small and big wide data. CEPA Working Paper No. WP09/2020. 17

Zelenyuk, N., Faff, R., and Pathan, S. (2019). Size-conditioned mandatory capital adequacy disclosure and bank intermediation. *Accounting & Finance*, doi:10.1111/acfi.12536. 6, 6.1, 6.2

Zelenyuk, N., Faff, R., and Pathan, S. (2020). The impact of voluntary capital adequacy disclosure on bank lending and liquidity creation. *Accounting & Finance*, forthcoming. 6, 6.4, 6.4

Zelenyuk, N. and Zelenyuk, V. (2014). Regional and ownership drivers of bank efficiency. Working Paper 11, Centre for Efficiency and Productivity Analysis. 2.3

Zelenyuk, V. (2020). Aggregation of inputs and outputs prior to data envelopment analysis under big data. *European Journal of Operational Research*, 282(1):172 – 187. 17