Title Page (with Author Details)

A Principal-Agent Approach for Estimating Firm Efficiency: Revealing Bank Managerial Behavior

Levent Kutlu[2,*] **Emmanuel Mamatzakis**[2] **Mike G. Tsionas**[3]

Abstract

We consider agency-theory-based structural and reduced form models of bank performance. In the structural models, we consider the managerial decision-making processes, and reveal underlying managerial efforts and thereby managerial behavior. In an empirical application, we estimate performance of Eurozone banks using our novel structural and reduced form approaches by Markov Chain Monte Carlo techniques. Our findings show, for the first time, that bank underperformance persists in the Eurozone whereas considerable variability across Member States exists. Our agency-theory-based structural modelling would favor cooperation of all interested parties and towards higher financial integration.

Keywords: Banking; Competition; Bayesian estimation; Governance; Moral Hazard; Sources of efficiency

Declaration of Interest: None

[1] Corresponding Author: Department of Economics and Finance, University of Texas Rio Grande Valley, TX 78539, USA; levent.kutlu@utrgv.edu

[2] Birkbeck College, University of London, Malet Street, Bloomsbury, London WC1E 7HX, UK. e.mamatzakis@bbk.ac.uk

[3] Montpellier Business School, 2300 Avenue des Moulins, 34080 Montpellier, France. Lancaster University Management School, LA1 4YX, U.K., m.tsionas@lancaster.ac.uk

A Principal-Agent Approach for Estimating Firm Efficiency: Revealing Bank Managerial Behavior

Abstract

We consider agency-theory-based structural and reduced form models of bank performance. In the structural models, we take into account the managerial decision-making processes, and reveal underlying managerial efforts and thereby managerial behavior. In an empirical application, we estimate performance of Eurozone banks using our novel structural and reduced form approaches by Markov Chain Monte Carlo techniques. Our findings show, for the first time, that bank underperformance persists in the Eurozone whereas considerable variability across Member States exists. Our agency-theory-based structural modelling would favor cooperation of all interested parties and towards higher financial integration.

Keywords: Banking; Competition; Bayesian estimation; Governance; Moral Hazard; Sources of efficiency

1 Introduction

Corporate governance is concerned with the mechanisms by which the stakeholders of a firm exercise control over the management.[1] The managers (including corporate insiders) control the key decisions of the firm. However, the objectives of managers may not align with the objectives of the stakeholders, which leads to a principal agent problem.[2] The stakeholders in corporations would like to keep managers responsive to their objectives. One potential way to achieve this objective is having independent directors. Maug (1997) examines the role of independent directors from an optimal contracting perspective. He finds that unless the information is too costly, directors are the optimal institutions to check managerial discretion. Among others, an alternative mechanism that can affect the interplay between stakeholders and management is the competitiveness of the relevant market. In this study, we aim to understand how competition and managerial behavior would affect the success of stakeholders in terms of keeping the objectives of managers aligned with those of the stakeholders. We may proxy the extent of alignment through the (cost) efficiency of firm. In line with this, one of the important interests of corporate finance literature is examining the interplay between firm performance and governance.

The firm performance can be measured in a variety of ways. In the corporate finance literature, financial ratios such as return on equity and return on assets have been used dominantly (e.g., McConnell and Servaes, 1990; Thomsen and Pedersen, 2000; Demsetz and Villalonga, 2001; Claessens et al., 2002; Tian and

[1] We will use firm and corporation interchangeably.

[2] See John and Senbet (1998) for a study that surveys the empirical and theoretical literature on corporate governance mechanisms and related agency problems.

Estrin, 2008). However, financial ratios can be manipulated for tax or other purposes (Schulze et al., 2001; Durand and Vargas, 2003; Frydman and Jenter, 2010). Alternatively, firm performance may be measured by benchmarking methods that measure the performance of a firm via the radial distance from the frontier (e.g., cost frontier, profit frontier, production frontier.). The DEA and stochastic frontier analysis (SFA) are two popular approaches that are used for estimating firm efficiency. However, both DEA and SFA are reduced form approaches. In contrast to these reduced form approaches, we present a novel behavioral approach, which considers both competition aspects and the conflict of interest between stakeholders and managers. In particular, we consider a principal-agent framework, which we use for estimating efficiencies of firms. Moreover, in this context, we examine the theoretical relationship between firm efficiency and number of firms, which is used as a proxy for extent of competition. Hence, the relationship between firm competition and efficiency is one of the interests of our study.

The relationship between firm characteristics, governance, and efficiency are examined by many studies including: Kutlu and McCarthy (2016) (state-owned vs private-owned and efficiency); Bottasso and Sembenelli (2004), and Hanousek et al. (2012) (foreign ownership and efficiency); Margaritis and Psillaki (2007) and Weill (2008) (capital structure and efficiency); Almanidis et al. (2019) (firm size and efficiency); Baek and Pagán (2002) and Kutlu and Nair-Reichert (2020) (executive pay and efficiency); Koetter et al. (2012)(competition and efficiency).

In this study, we aim to identify the key managerial behavior that is related to firm inefficiency via game theoretical models and propose novel estimation methodologies. For this purpose, we provide two agency-theory-based structural models (see Berr, 2011), which differ by timing of choice variables.[3] In the first model, the managers play a two-stage game in which the managers first choose inefficiency (or lack of managerial effort) levels and, in the second stage, they compete on quantity. In the second model, the managers play a simultaneous game in which the managers choose inefficiency and quantity simultaneously. Although a manager cares about the profit of firm, she also cares about her effort level. Hence, the objective of the manager is not fully aligned (see Francis et al., 2015) with profit maximization. This leads to a deviation from the optimal profit level. These models illustrate that the equilibrium inefficiencies of firms depend on the number of competing firms, the managers' utility function from lack of effort, the terms of contract, and the "average inefficiency." Also, we illustrate that the standard non-structural stochastic frontier models suffer from a serious misspecification issue as these models ignore the structural equations that describe the managerial behavior and market competition. This misspecification issue can be recast as an endogeneity problem that results from an omitted variable in the cost function. Therefore, conventional studies that examine the interplay between firm efficiency and governance of a firm, may give seriously flawed results.

We decompose the firm inefficiency into two parts: structural and non-structural.[4]

[3] See Ceccini et al. (2013) and Assaf et al. (2019) for studies that develop methods for solving principal-agent problems.

[4] See Gathon and Pestieau (1995) for a decomposition of productive efficiency into a management and a regulatory component. See Muñiz (2002) for a paper on separating managerial inefficiency and external conditions. Finally, see Berger and DeYoung (1997) and Williams (2004) for studies that examine causal relationships between problematic loans, cost efficiency, and bank capital. They consider 4 hypothesis: 1) Bad luck, 2) Bad management, 3) Skimming, 4) Moral hazard. Under bad luck hypothesis, external events precipitate an increase in problem loans for the bank. They expect increases in nonperforming loans to Granger-cause decreases in measured cost efficiency. In our paper, we interpret bad luck as anything that is uncontrollable by managers that causes reduction in cost efficiency. We measure this by what we call unstructural inefficiency. While their moral hazard hypothesis relates to excessive risk-taking behavior when another party is bearing part of the risk and cannot easily prevent risk-taking; our principal agent

The structural component is the part of the inefficiency that is occurring by the conscious will of the bank manager. The non-structural inefficiency, however, is the inefficiency that occurs, supposedly, due to unplanned factors such as mistakes by the bank managers and other units in the firm or environmental factors. Basically, the non-structural inefficiency is the part of inefficiency that is not explained by our structural model. This decomposition allows us to have some idea about the extent to which we have control over inefficiency. For example, it would be easier to control the structural inefficiency by improving contracts and choosing the right managers but for the non-structural inefficiency these tools may not be as effective. In the literature there are studies that proposed a variety of decompositions to efficiency. For instance, Gathon and Pestieau (1995) decompose productive efficiency into a management and a regulatory component; Muñiz (2002) separate managerial inefficiency and external conditions; and Berger and DeYoung (1997) and Williams (2004) examine causal relationships between problematic loans, cost efficiency, and bank capital. Berger and DeYoung (1997) and Williams (2004) consider four hypothesis: 1) Bad luck, 2) Bad management, 3) Skimming, 4) Moral hazard. Under bad luck hypothesis, external events precipitate an increase in problem loans for the bank. They expect increases in nonperforming loans to Granger-cause decreases in measured cost efficiency. In our paper, we interpret bad luck as anything that is uncontrollable by managers that leads to reduction in cost efficiency. We measure this by what we call non-structural inefficiency. While their moral hazard hypothesis relates to excessive risk-taking behavior when an-

problem relates to cost minimization effort of manager based on exogenously given contract. We control risk using two types of control variables in our model, i.e., z-score and ratio of equity to assets.

other party is bearing part of the risk and cannot easily prevent risk-taking; our principal agent problem relates to cost minimization effort of the manager based on exogenously given contract. We control risk using two types of control variables in our model, i.e., z-score and ratio of equity to assets.

Among others, empirical work on the firm efficiency has favored the banking sector (see Lozano-Vivas, 1997; Paradi et al., 2012; Galan et al., 2015; Dong et al., 2016; Delis et al., 2017)[5] since the success of banks takes a central role in overall health of the economy; and thus understanding the determinants of bank efficiency is essential. We apply our structural and reduced form approaches to the Eurozone banks for the time period between 2002 and 2015. The banking industry of the Eurozone is of interest in particular due to the persistence of banking imbalances well after the financial crisis in 2008, and despite the common currency zone. Joyce et al. (2012) highlight these banking imbalances in the Eurozone, by going as far as suggesting the situation with banking imbalances has been rather severe in the case of the Eurozone that could resemble kind of a bank run as deposits from the periphery have been directed to the core of the Eurozone. Sinn and Wollmershauser (2011) discuss also this and argue that there might exist 'target system imbalances' that could impair bank performance, mainly in the periphery of the Eurozone. In the empirical section, we focus on this prior evidence and consider the periphery of the Eurozone vs the core of the Eurozone. This exercise would allow us to identify any shifts in structural vs non-structural inefficiency across Member States in the Eurozone and could reveal any asymmetries between the core and the periphery.

Our novel stochastic frontier approaches call for certain Markov Chain Monte

[5] See Tsionas (2007) for a paper on microfoundations of stochastic frontier analysis.

Carlo (MCMC) techniques to implement Bayesian inference. The reduced form model assumes that inefficiency of a bank may be affected by the average of other banks' inefficiencies in the same market. The Bayes factor comparison of these models favors the two-stage structural model. Hence, in our banking context, the managers seem to have a two-stage decision scheme where first they determine the inefficiency and then the quantity. It appears that, the managers plan their inefficiency levels ahead even before competing in the market. Our results show that, the inefficiency that is planned by the managers (i.e., structural inefficiency) is less than the unplanned inefficiency. More interestingly, while non-structural inefficiencies are somewhat similar, both structural and non-structural inefficiencies are higher for the periphery of Eurozone than the core of the Eurozone.[6]

In what follows, the next section presents our theoretical model, whilst section 3 presents the data and reports the empirical application. Lastly, section 4 offers concluding remarks.

2 Model and Theoretical Results

Stochastic frontier models assume a composed error term. The first component is the usual two-sided error term and the second component is a one-sided error term, which aims to capture inefficiency.[7] Although fully efficient production is the standard in neoclassical economics, in practice, we find that most firms are

[6] For $2008-2015$ the non-structural inefficiency is slightly higher for the core countries. However, overall, non-structural inefficiency is higher for periphery countries.

[7] A variety of distributions is proposed for the one-sided error component. For example, Aigner et al. (1977) use the half normal; Meeusen and van den Broeck (1977) use the exponential; Stevenson(1980) use the truncated normal; Greene (1980a, 1980b, 2003) use the gamma distributions.

away from the frontier.[8] To the extent that increasing efficiency involves costs, utilization of real resources as well as managerial skill, it is natural to think that in order to motivate efficiency, we need a formal model of the firm. Specifically, we need an agency model that accounts for the fact that managerial effort has to be motivated and remunerated. Simply assuming all these factors away to arrive at a stochastic frontier model with two error components is not enough if we need a solid micro-foundation for efficiency.

Structural stochastic frontier models are not common in the literature. The closest study to ours seems to be that of Gagnepain and Ivaldi (2002), which estimates a structural model of efficiency when the firms are regulated. Similar to our study, they consider a model that is based on the agency theory. One of the key differences of our model is that we incorporate competition to our agency problem. Hence, we consider the game theoretic interactions of the firms which affect the inefficiency. The connection between the extent of competition and inefficiency has been acknowledged by strands of studies (e.g., Hicks, 1935; Berger and Hannan, 1998; Maudos and de Guevara, 2007; Delis and Tsionas, 2009; Koetter et al., 2012; Kutlu and Sickles, 2012; Tsionas et al., 2018). A structural stochastic model that is ignoring this relationship appears to be missing a vital determinant of inefficiency. Hence, among other contributions, our study fills this gap. Kutlu and Wang (2018a) propose another structural stochastic frontier model that estimates inefficiency in a conduct parameter game.[9] Hence, their

[8]While inefficient productive units are frequently observed, there are occasions where many firms may be fully efficient. For example, Green and Mayes (1991) report that over 30% of samples of UK and Australian industries had the "wrong" skewness, which may suggest full efficiency for these samples.

[9]See Corts (1999) and Kutlu and Sickles (2012) for more details about conduct parameter (or conjectural variations) game. See also Kutlu and Wang (2018b) for an application of Kutlu and Wang (2018a).

model incorporates competition when estimating inefficiency. However, they only estimate what we call non-structural inefficiency although their model is structural. That is, they model how the firms compete in the market but keep inefficiency term as a one-sided random variable, which ignores the structural agency problem. Martin (1993) considers a theoretical model of principal-agent problem where the firms compete in a Cournot setting.[10] He examines the relationship between the number of firms and efficiency. In order to derive general theoretical results, this study assumes a linear demand function and a particular functional form for the cost function, which restricts the applicability of this model in general empirical settings.

The purpose of this section is presenting our models and results, which would help understanding the sources of firm inefficiency as well as proposing ways to estimate inefficiency. We examine this issue by presenting two game theoretic models where the managers choose their effort levels in an oligopoly setting. In our framework, the inefficiency is realized because the objectives of the shareholders (i.e., profit maximization) and manager (i.e., utility maximization) are not fully aligned. We assume that the shareholders of firm i care about the profit:

$$\pi_i(\mathbf{q}, \mathbf{r}) = P(Q(\mathbf{r}))\, q_i(\mathbf{r}) - C_i^*(q_i, r_i), \tag{1}$$

where q_i is the output quantity for firm i, $Q = \sum_i q_i$ is the total quantity, $r_i \geq 1$ is a variable that represents the lack of effort by manager i, $\mathbf{r} = (r_1, r_2, ..., r_n)'$, $\mathbf{q} = (q_1, q_2, ..., q_n)'$, and $P(Q(\mathbf{r}))$ is the inverse demand function so that $P'(Q(\mathbf{r})) < 0$. In line with the stochastic frontier literature, we assume that the total cost function

[10] For a related study see also Krishna (2001).

is separable so that $C_i^*(q_i, r_i) = C_i(q_i) r_i$, where $e_i = 1/r_i \in [0, 1]$ is defined as the efficiency. The utility function for manager i is given by:[11]

$$\tilde{U}_i(\mathbf{q}, \mathbf{r}) = a_i \pi_i(\mathbf{q}, \mathbf{r}) + \tilde{V}_i(r_i) + A_i, \tag{2}$$

where $\tilde{V}_i(r_i) \geq 0$ is a function that represents the utility of manager from lack of effort; $a_i \in (0, 1)$ is a constant that represents the portion of the income of manager that depends on the profit; and A_i is the fixed salary. Among others $\tilde{V}_i$ may be determined by the manager's age, proximity to retirement, education background, talent, and managerial power as well as cultural factors. For example, Kauko (2009) finds that manager's age and education have strong effects on efficiency.

After normalization, without loss of generality, the utility function becomes:[12]

[11] In general, the utility of the manager may be a non-linear function of π_i, $\tilde{V}_i$, and A_i. In order to simplify analysis, we assume a quasi-linear form, which is commonly used in the principal-agent literature. One important property of quasi-linear utility form is that Marshallian/Walrasian demand for lack of effort does not depend on wealth.

[12] We assume that $V_i'(r_i) > 0$ and $V_i''(r_i) < 0$. Hence, the (normalized) utility from lack of effort is an increasing function. The second order condition implies that increase in lack of effort provides less additional satisfaction to the manager. In practice, V_i, and thus equilibrium inefficiency, depends on observed or unobserved characteristics of the managers and the terms of contract, i.e., a_i and A_i. Technically $V_i''(r_i) < 0$ assumption resembles to the monotonicity assumptions of OP-LP models, which enable the invertibility of a relevant function. The monotonicity assumptions of OP-LP may not be applicable in markets with some market power. For example, LP assumes perfect competition to assure monotonicity. In contrast to models of OP-LP, our model allows and works when the firms have market power. Since the presence of market power is not rare, we believe that it is essential to have a structural model that allows for market power. In our case, the monotonicity assumption, $V_i''(r_i) < 0$, is a standard concavity assumption for utility functions.

[13]

$$U_i(\mathbf{q}, \mathbf{r}) = \pi_i(\mathbf{q}, \mathbf{r}) + V_i(r_i). \tag{3}$$

We consider two scenarios. The first scenario models r_i via a two-stage game where in the first stage the managers decide about their effort levels. In the second stage, conditional on their predetermined effort levels, they play a quantity choice game where the payoff functions are the utilities of managers. This scenario may be reasonable when the salaries of the managers are determined around the beginning of a fiscal year. Based on the salary and bonuses, the managers may choose their effort levels for the upcoming year. Moreover, whenever possible, personal travel and vacation plans may be made well ahead of fiscal year, which affect the work performance.[14] In the second scenario, the managers decide about their effort levels and quantities simultaneously.

[13] The relationship between competition and efficiency has been analyzed by a plethora of papers (e.g., Hicks, 1935; Berger and Hannan, 1998; Maudos and de Guevara, 2007; Delis and Tsionas, 2009; Koetter et al., 2012; Kutlu and Sickles, 2012; Andrieş and Căpraru, 2014; and Tsionas et al., 2018). Some of these studies use reduced form analysis and others use the conduct parameter approach. We extend the literature and we propose a new approach whereby the relationship between competition and efficiency is examined through a novel principal agent model. Unlike our model, most principal agent studies do not incorporate agent based competition. Hence, our theoretical model combines competition, principal agent problem, and efficiency. The main focus of our proposed modelling approach is to decompose and estimate structural and nonstructural efficiency for the first time to the best of our knowledge. The importance of such a decomposition is that we offer an understanding to what extent efficiency can be controlled through policies that affect competition or managerial compensation. For example, suboptimal managerial payment schedules may cause structural inefficiency and we believe that understanding the magnitude of such inefficiency relative to nonstructural inefficiency is a relevant question.

[14] An alternative two-stage game, which we do not consider, may be where in the first stage the managers decide about their effort levels and, in the second stage, they decide about their effort levels.

2.1 The Structural Games

We start with the two-stage game where the managers decide about their efficiency and quantity choices sequentially. Following the standard approach, we solve the equilibrium of the game backwards. First, we examine the solution of second stage conditional on $\mathbf{r}$. In the second stage, the optimization problem of manager i is:

$$\max_{q_i} U_i(\mathbf{q}, \mathbf{r}). \tag{4}$$

Given $\mathbf{r}$, the optimization problem of the managers for the second stage is given by:

$$\frac{\partial U_i(\mathbf{q}, \mathbf{r})}{\partial q_i} = P'(Q(\mathbf{r}))\, q_i(\mathbf{r}) + P(Q(\mathbf{r})) - C_i'(q_i)\, r_i = 0. \tag{5}$$

It turns out that the equilibrium profit of firm i depends not only on the lack of effort by manager i but also the average lack of effort by other managers, $\bar{r}$. In particular, $\bar{r} - r_i$ measures the extent to which firm i is farther away from the "mean firm." As the other firms get better relative to firm i (i.e., $\bar{r}$ decreases), the profit of firm i gets lower. In turn, manager i's equilibrium decision for her lack of effort, r_i, critically depends on other managers' behavior. In the empirical section, we utilize this observation when we estimate a reduced form model as an alternative to the structural games that we present.

The equilibrium value for firm i, $q_i(\mathbf{r})$, critically depends on inefficiency levels of other managers. This is in line with the standard prediction of Cournot models that the firms with lower marginal costs produce more. We find that the difference from average market quantity is positively related with the difference between marginal cost of the average firm and marginal cost of firm i; and negatively related with the (absolute value of) slope of the inverse demand function.[15] Hence, when the inverse demand function is very steep, most of the efficiency advantage is due to only higher price-marginal cost markup. However, if the inverse demand function is relatively flat, the firm not only benefits from price-marginal cost markup but also selling more relative to others.

We also find that q_i and Q decrease as the lack of effort, r_i, increases. In most cases, as the other managers get more inefficient, q_i increases.[16]

Although, higher inefficiency increases the utility of manager from lack of effort, the profit of firm decreases by higher inefficiency levels. Hence, the optimal inefficiency level of a manager is determined by these conflicting effects.[17]

Now, we consider the optimization problem of managers in the first stage where they choose the optimal level for lack of effort:

$$\max_{r_i} U_i\left(q\left(\mathbf{r}\right), \mathbf{r}\right). \tag{6}$$

Under the constant marginal cost assumption, (i.e., $C_i\left(q_i\right) = cq_i$) the first order conditions for the first stage become:

$$\begin{aligned} V_i'\left(r_i\right) &= \frac{\partial P'\left(Q\left(\mathbf{r}\right)\right) q_i^2\left(\mathbf{r}\right)}{\partial r_i} \\ &= q_i\left(\mathbf{r}\right) c \frac{2n + \theta\left(\mathbf{r}\right)\left(2 - s_i\left(\mathbf{r}\right)\right)}{n + 1 + \theta\left(\mathbf{r}\right)}. \end{aligned} \tag{7}$$

An alternative model would have been a simultaneous game in which the managers make their decisions about efficiency and quantity simultaneously. The corre-

[15] For mathematical formulation of these results, see Proposition 1 in Appendix A.
[16] For mathematical formulation of these results, see Proposition 2 in Appendix A.
[17] For mathematical formulation of these results, see Proposition 3 in Appendix A.

sponding econometric supply equations for the simultaneous setting can be written as follows:[18]

$$P_i = -P_i' q_i + C_i' V_i'^{-1}(C_i) + v_{1i} \tag{8}$$
$$\ln C_i^* = \ln C_i + \ln V_i'^{-1}(C_i) + u_i + v_{2i},$$

where v_{1i} and v_{2i} are the two-sided error terms, $u_i \geq 0$ is a one-sided error term that aims to capture non-structural inefficiency, and $\ln r_i = \ln V_i'^{-1}(C_i) \geq 0$ is the key component of our model that is representing the structural inefficiency.[19] Note that the theoretical model doesn't have the error terms v_{1i}, v_{2i}, and u_i. Here v_{1i} and v_{2i} are the usual error terms in econometric models; and u_i captures the unobserved non-structural inefficiency that is not captured by the structural inefficiency term. These expressions can be simultaneously estimated with the demand equation. Since $\ln V_i'^{-1}(C_i)$ is a function of C_i, the conventional stochastic cost frontier models will be misspecified as they assume $\ln V_i'^{-1}(C_i) = 0$. Hence, conventional corporate finance studies that aim to understand the relationship between governance and firm efficiency, may give misleading results due to such misspecification as such models do not properly incorporate game theoretical interactions between managers. We turn to this issue in our empirical application.

While the structural inefficiency is fully determined by the manager given the contract, it is possible that the contract may be suboptimal. Our structural in-

[18] The first order conditions for the simultaneous setting are given in Appendix A.

[19] A common choice for the distribution for u_i would be the half-normal or truncated normal distributions. Another choice would be using the distribution free approach of Cornwell et al. (1990). Moreover, note that in our model we assume that the structural and nonstructural components of inefficiency enter the model additively. Other modeling choices such as multiplicative decomposition would be possible but since the model is already given in logarithm form, we prefer an additive decomposition.

efficiency term, $\ln V_i'^{-1}(C_i)$, captures the inefficiency conditional on the observed contract. Hence, it is possible to affect the structural efficiency by improving the existing contracts. Somewhat similarly, in their agency based structural model, Gagnepain and Ivaldi (2002) do not assume that actual regulatory mechanisms are optimal. So, they take their characteristics as given for estimating the model parameters, which implicitly assumes that the choice of regulation is exogenous.

3 Empirical Example: Eurozone Banking Industry

A large part of efficiency literature is focusing on the European banking systems (e.g. Allen and Rai, 1996; Altunbas et al., 2001; Lozano-Vivas et al., 2002; Maudos et al., 2002; Brissimis et al., 2010; and Mamatzakis et al., 2015).[20][21] We examine the Eurozone bank inefficiencies using our structural and reduced form approaches. As described earlier, our structural models build upon the agency theory where the managers endogenously choose their effort levels based on their heterogeneous utilities, which may depend on their ability levels. Sherman and Gold (1985), one of the first applications of data envelopment analysis (DEA) to banking, use DEA on 14 branches of one bank and report that "four of the inefficient branches were believed to be run by weak managers." They also find that smaller branches tended to be less efficient, which is in line with the view that the management is likely

[20] See Bikker, J.A. and Bos, J.W. (2008) for a good coverage of theoretical and empirical analysis of bank efficiency.

[21] Another popular country for banking studies is the US. See, for example, Delis et al. (2017) and Glass, A.J. and Kenjegalieva, K. (2019) for recent papers that study the US bank efficiency.

to assign less experienced managers to the small branches. Hence, while they find a connection between managerial quality and inefficiency, their method is non-structural and they do not differentiate non-structural and structural inefficiency.

3.1 Data

The data set includes commercial, cooperative, savings, investment, and real-estate banks in Eurozone countries that are listed in the IBCA-Bankscope database over the period $2002 - 2015$. After reviewing the data for reporting errors and other inconsistencies, we obtain an unbalanced panel dataset that consists of banks in 14 countries of the Eurozone. All bank-specific variables are obtained from Bankscope, at annual frequency, and in thousand USD. There are $23,917$ observations of $3,229$ banks.

The periphery of the Eurozone includes 6 countries, namely Cyprus, Greece, Ireland, Italy, Portugal, and Spain. The main common characteristic of all the periphery is that it has received financial assistance from the EU and the IMF due to the aftermath of the financial crisis that led into a banking crisis and to a public debt crisis, due to the capitalization that followed. The remaining Eurozone countries would form the core group, namely Austria, Belgium, Finland, France, Germany, Luxembourg, Malta, and the Netherlands. For consistency in the underlying data generating process and due to limitations in number of observations, we exclude Estonia, Latvia, Lithuania, Slovenia, and Slovakia from the sample as these Member States joined the single currency later than 2007.

For the definition of bank inputs and outputs, we follow the vast majority of the literature and employ the financial intermediation approach[22] proposed by Sealey and Lindley (1977), which assumes that the bank collects funds and using labor and physical capital transforms them into loans and other earning assets. In order to address issues related to outliers in the underlying data generating process, our sample includes banks that report positive equity capital. In addition, we carefully review the data for reporting errors and other inconsistencies. In some detail, we exclude banks for which bank inputs and outputs are outside the ± 3 standard deviation interval.

In particular, we specify three inputs: capital, labor, and deposits; and two outputs: loans and other earning assets (which include government securities, bonds, equity investments, CDs, T-bills, and equity investment). The price of labor is defined as the ratio of personnel expenses to number of employees. The price of capital is defined as the ratio of other administrative expenses to fixed assets. The price of deposits is the ratio of total interest expenses to total borrowed funds. For the calculation of output prices, the price of loans is defined as the ratio of interest income to total loans, while the price of other earning assets is defined as total non-interest income to total other earning assets.

For modelling cost function, we used two measures of risk: i) the bank specific z-score[23] (see Delis et al., 2017) as well as the country-specific z-score, which is calculated as a weighted average where the weights being the share of each bank's total assets to respective country's total assets ii) the ratio of equity to assets. Moreover, we include dummy variables for commercial, investment, real-estate,

[22] For a review of the various approaches that have been proposed in the literature for the definition of bank inputs and outputs see Berger and Humphrey (1997).

[23] The z-score is defined as follows: $Z = (ROA + CR)/\sigma(ROA)$, where ROA is the return on assets, CR is the capital ratio, and $\sigma(ROA)$ is the standard deviation of ROA.

and savings banks.[24] The financial crisis of 2008 had a major impact on individual investors' wealth (Hudomiet et al., 2011; Hoffmann and Post, 2013). In order to control for possible changes in identified factors (managers characteristics, market power, heterogeneity) caused by 2008 crisis, we also include a 2008-2015 dummy (See Fujii et al., 2017) and its interactions with all control variables.[25] Besides this, time fixed effects and country fixed effects are controlled as well. Moreover, when announcing our parameter estimates, we present separate posterior mean estimates for pre- and post-crisis.

The additions to the sample are not necessarily new market entrants, but rather successful banks that are added to the database over time. Exits from the sample are due either to bank failures or to mergers with other banks or are a consequence of changes in the coverage of the Bankscope database. Our sample covers the largest credit institutions in each country, as defined by their balance sheet aggregates. Due to the specific features of the German banking system (large number of relatively small banks), our sample is dominated by German banks.[26]

[24] In estimations, we drop cooperative banks dummy variable.

[25] Among others, Pi and Timme (1993) have identified principal-agent based issues, which affect decision management and decision control processes with an end result to efficiency. In addition, a performance-ownership link has been well identified in the previous literature (Demsetz and Lehn, 1985; Fuentes and Vergara, 2007). However, we cannot control ownership structure of banks due to data availability issues. Nevertheless, the relationship between ownership structure and efficiency could be accommodated in our proposed modelling approach by including ownership related control variables when modelling both structural and non-structural efficiency components though it is not the main emphasis of the present paper. Also note that efficiency is captured by a random variable that is bank-specific, which may help capturing efficiency in a flexibly manner.

[26] Our sample considers the size of balance sheet per bank and per country. For example, due to specific features of the German banking system (large number of relatively small banks), we take into account in the empirical application the underlying total assets per bank. This corrects for the large number of German banks of any EU bank sample. Note that although, the number of banks from Germany would be above the corresponding figure of other countries, for example France, focusing on the number of banks might be misleading as bank' total assets are of importance for the size of the country banking industry. To this end, banking industries in countries like France and UK would weight higher than the German banking industry as total

Descriptive statistics of the data are provided in Table 1.[27]

Table 1.a. Descriptive Statistics for the Eurozone

VARIABLE	MEAN	STD. DEV.	MIN	MAX
C*: Total Cost	104673.9	150059	5150	478000
W1: Capital Price	1.261735	1.168332	0.396904	4.24958
W2: Labor Price	1.587868	1.194612	0.540723	4.49794
W3: Deposit Price	0.0187389	0.0077274	0.008237	0.031838
Q1: Loans	1594909	2396045	62202	7700000
Q2: Other Earning Assets	899574	1291166	42624	4100000
Total Assets	2921072	4356429	129103	1.40E+07
Equity	200402.6	282850.8	11829	913049
Price of Loans	0.052731	0.0120246	0.0355171	0.0729045
Price of Other Earning Assets	0.0375926	0.0225455	0.01573	0.0910299
Z-score_B: Bank Specific	4.124618	3.602281	0.2508916	11.59291
Z-score _C: Country Specific	3.088586	3.277355	0.2508916	11.59291
Equity to Assets	0.0832904	0.0288753	0.0458019	0.1384615
Commercial Bank Dummy	0.1912706	0.3933099	0	1
Investment Bank Dummy	0.5236771	0.4994496	0	1
Real-Estate Dummy	0.0270054	0.1621025	0	1
Savings Bank Dummy	0.0315554	0.174817	0	1
Cooperative Bank Dummy	0.2264914	0.4185695	0	1

Note: The table presents descriptive statistics of the main variables used in our analysis for the full sample over the examined period, 2002-2015. Number of observations 23,615 and the Eurozone includes the countries: Austria, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, the Netherlands, Portugal, and Spain.

Table 1.b. Descriptive Statistics for the Periphery

VARIABLE	MEAN	STD. DEV.	MIN	MAX
C*: Total Cost	88486.06	139389.2	5150	478000
W1: Capital Price	1.087192	1.089453	0.396904	4.24958
W2: Labor Price	1.326478	1.140832	0.540723	4.49794
W3: Deposit Price	0.0175173	0.0073033	0.008237	0.031838
Q1: Loans	1659343	2574135	62202	7700000
Q2: Other Earning Assets	831208	1344284	42624	4100000
Total Assets	3006790	4718866	129103	1.40E+07
Equity	216185.3	302880.3	11829	913049
Price of Loans	0.0485132	0.011877	0.0355171	0.0729045
Price of Other Earning Assets	0.0401392	0.0222235	0.01573	0.0910299
Z-score_B: Bank Specific	1.763584	1.763424	0.2508916	11.59291
Z-score _C: Country Specific	1.665594	1.477467	0.2508916	11.59291
Equity to Assets	0.0979033	0.0296494	0.0458019	0.1384615
Commercial Bank Dummy	0.1788589	0.3832677	0	1
Investment Bank Dummy	0.6435072	0.4790051	0	1
Real-Estate Dummy	0.039552	0.1949212	0	1
Savings Bank Dummy	0.0122506	0.1100121	0	1
Cooperative Bank Dummy	0.1258313	0.3316882	0	1

Note: The table presents descriptive statistics of the main variables used in our analysis for the full sample over the examined period, 2002-2015. Number of observations 5,712 and the periphery of the Eurozone includes the countries: Cyprus, Greece, Ireland, Italy, Portugal, and Spain.

---

assets in France per bank are larger than in Germany. In our sample, to deal with the large number of German small banks, we consider banks' balance sheet size that is total assets as a control variable. In addition, in the empirical analysis we control for country fixed effects.

[27] Number of banks per country are as follows: Austria 211, Belgium 32, Cyprus 16, Finland 33, France 224, Germany 1489, Greece 10, Ireland 17, Italy 506, Luxembourg 60, Malta 10, Netherlands 31, Portugal 99, and Spain 109.

Table 1.c. Descriptive Statistics for the Core

VARIABLE	MEAN	STD. DEV.	MIN	MAX
C*: Total Cost	109787.1	152902.2	5150	478000
W1: Capital Price	1.31756	1.187085	0.396904	4.24958
W2: Labor Price	1.672114	1.19939	0.540723	4.49794
W3: Deposit Price	0.0191286	0.007818	0.008237	0.031838
Q1: Loans	1574378	2336153	62202	7700000
Q2: Other Earning Assets	921357.7	1273043	42624	4100000
Total Assets	2893895	4234822	129103	1.40E+07
Equity	195398.6	276017.7	11829	913049
Price of Loans	0.0541234	0.0117452	0.0355171	0.0729045
Price of Other Earning Assets	0.0367811	0.0225875	0.01573	0.0910299
Z-score_B: Bank Specific	4.873201	3.711761	0.2508916	11.59291
Z-score _C: Country Specific	3.646219	2.973446	0.2508916	11.59291
Equity to Assets	0.0786572	0.027024	0.0458019	0.1384615
Commercial Bank Dummy	0.1952059	0.3963701	0	1
Investment Bank Dummy	0.4856842	0.4998089	0	1
Real-Estate Dummy	0.0230274	0.1499947	0	1
Savings Bank Dummy	0.0376762	0.1904172	0	1
Cooperative Bank Dummy	0.2584064	0.4377707	0	1

Note: The table presents descriptive statistics of the main variables used in our analysis for the full sample over the examined period, 2002-2015. Number of observations 18,022 and the core of the Eurozone includes the countries: Austria, Belgium , Finland, France, Germany, Lux embourg, Malta, and the Netherlands.

3.2 Empirical Models and Results

In this section, we provide the details of our agency-theory-based models that we use for estimating the efficiencies of Eurozone banks. We use loans as Q_1 and other earning assets as Q_2. The inputs to production are capital, labor, and deposits, which comprise total costs. The cost function is assumed to have a translog functional form of outputs, input prices, and total assets. We include total assets as a bank size measure as it may indicate higher diversification of a bank's loan portfolio (Mester, 1993), where logarithm of total assets is used to control for different bank sizes.

Besides these variables, as mentioned in the data section, our cost function includes bank specific and country specific z-score as well as the ratio of equity to assets. The variables are interacted with all other variables in the cost function and the squares are included as well. We also include a $2008-2015$ dummy and its full interactions with all other variables for the cost function variables.[28] This

[28] The $2008-2015$ dummy and its cross-products are jointly significant at any conventional significance level.

dummy and its interactions is capturing the financial meltdown of 2008 that could have contributed to a major structural break in the underlying fundamentals of the banking industry in the Eurozone that is not captured by the technology. Although Joyce et al. (2012) argue that the Eurozone was rather resilient early on to the financial crisis originated in the US, by 2011 the crisis had a major detrimental impact to the Eurozone. Besides these we control for country and time fixed effects by dummy variables.

As Eurozone countries constitute a single currency union with some noticeable heterogeneity in their banking industries, we estimate bank inefficiency for the three main groups: the Eurozone, the periphery, and the core of the Eurozone. Heterogeneity is addressed by including country fixed effects. As discussed, it could be the case that the Eurozone is rather heterogenous despite representing a single currency area. Indeed, bank imbalances in the aftermath of the financial crisis might have contributed in to two banking zones within the Eurozone; that is the periphery from where deposits fled due to poor performance to the core of the Eurozone (see Joyce et al. 2012).

In the estimations, we impose the homogeneity restriction to the cost function and other parametric restrictions for the V (the utility function for lack of effort) function as discussed in the theoretical section. We assume the following functional forms in our empirical model:[29]

$$
\begin{aligned}
\ln P_1(Q) &= \alpha_0 - \alpha_1 \ln Q_1 - \alpha_2 \ln Q_2 + X'\alpha \\
\ln P_2(Q) &= \beta_0 - \beta_1 \ln Q_1 - \beta_2 \ln Q_2 + X'\beta \\
C_i^*(q_i, r_i) &= C(q_i)\, r_i \\
V_i(r_i) &= \gamma_0 r_i^{\gamma_1},\ \gamma_0 > 0 \text{ and } 0 < \gamma_1 < 1,
\end{aligned}
\tag{9}
$$

where X is a vector of control variables for the demand equations. The parameters of the model are restricted so that curvature conditions for the (normalized) utility function for lack of effort hold, i.e., $V_i'(r_i) > 0$ and $V_i''(r_i) < 0$. In other words, the (normalized) utility from lack of effort is an increasing function and increase in lack of effort provides less additional satisfaction to the manager. The inverse of $V_i'(r_i)$ is essential for our model and is given by:

$$
\begin{aligned}
V_i'^{-1}(x) &= (\gamma_0\gamma_1)^{\frac{1}{1-\gamma_1}}\, x^{\frac{-1}{1-\gamma_1}} \\
&= \theta_0 x^{\theta_1} \geq 1,\ \theta_0 > 0 \text{ and } \theta_1 < -1.
\end{aligned}
\tag{10}
$$

3.2.1 Estimation of Structural Games

The two-stage structural game can be estimated using Equation (24) and Equation (25). The relevant supply equations for the simultaneous structural game are given

[29] Two-output model details are given in Appendix B.

by:[30]

$$
\begin{aligned}
P_{1it} &= \alpha_1 \frac{P_{1it}q_{1it}}{Q_{1t}} + \beta_1 \frac{P_2 q_{2it}}{Q_{1t}} + \frac{\partial C_{it}}{\partial q_{1i}} \theta_0 C_{it}^{\theta_1} + v_{1it} \\
P_{2it} &= \alpha_2 \frac{P_{1it}q_{1it}}{Q_{2t}} + \beta_2 \frac{P_{2it} q_{2it}}{Q_{2t}} + \frac{\partial C_{it}}{\partial q_{2it}} \theta_0 C_{it}^{\theta_1} + v_{2it} \\
\ln C_{it}^{*} &= \ln C_{it} + \ln r_{it} + u_{it} + v_{3it} \\
&= \ln \theta_0 + (\theta_1 + 1) \ln C_{it} + u_{it} + v_{3it},
\end{aligned}
\tag{11}
$$

where $\ln r_{it} = \ln \theta_0 + \theta_1 \ln C_{it} \geq 0$ is the term that captures the structural inefficiency and $u_{it} \geq 0$ term is a one-sided random variable added to capture the non-structural inefficiency. The standard stochastic frontier models would only estimate the cost function and assume that there is no structural inefficiency so that $\ln r_{it} = 0$, i.e., $\theta_0 = 1$ and $\theta_1 = 0$. Hence, conventional corporate finance studies that use standard stochastic frontier models, may have specification issues if the managers are acting strategically.

For the two-stage structural game, we use Equation (24) and (25) to get the parameter estimates.[31] For the simultaneous structural game, we estimate the System (11), which is a nonlinear simultaneous equations model with a non-trivial Jacobian term that is computed numerically.[32]

[30] After adding the error terms, we can use bank specific output prices.

[31] These first order conditions are given in Appendix A.

[32] The estimation details for the simultaneous structural game is given in Appendix C-E.

3.2.2 Reduced Form Model

Based on Equation (15), we argue that the equilibrium profit of firm i depends on the inefficiency of firm i as well as the "average inefficiency" of other firms.[33] Our reduced form model is in line with this idea. Hence, our model can be considered as some form of a spatial efficiency model. For notational simplicity, we illustrate the single market case. However, it is trivial to allow separate markets where the inefficiency of a firm in one market is not effected by the efficiencies of firms from other markets. Suppose we have a model of the form:

$$\begin{aligned} \ln C_{it}^{*} &= \ln C_{it}(\beta) + \varepsilon_{it} \\ \varepsilon_{it} &= u_{it} + v_{it} \\ u_{it} | \{u_{jt}, j \neq i\} &\sim \mathcal{N}_{+}\left(z_{it}'\gamma_1 + \gamma_2 u_{-it}, \sigma_u^2\right), \end{aligned} \tag{12}$$

where $u_{-it} = (n-1)^{-1} \sum_{j=1, j\neq i}^{n-1} u_{jt}$ along with the assumption that, independently, $v_{it} \sim$ i.i.d.$\mathcal{N}(0, \sigma_v^2)$. Here, u_{it} is a one-sided error term that is capturing inefficiency. We assume z_{it} is $m \times 1$ and includes a constant term. We consider u_{it}'s as latent variables, which are considered as parameters with appropriate priors.[34] For this procedure, knowledge of conditional distributions of parameters is required. We use non-informative priors. Also, we use $15,000$ iterations the first $5,000$ of which are discarded to mitigate possible start-up and convergence effects.[35]

[33] This is consistent with the social comparison theory of Festinger (1954) as well.
[34] See Tanner and Wong (1987) and Tsionas (2002) for details.
[35] We present the conditional distributions and details in Appendix C-E.

3.2.3 Results

In this section, we present our estimation results for our models. In output tables, the sample means of posterior mean estimates of parameters for a variety of percentiles based on total assets and standard deviations (in parenthesis) for the estimated distributions of the model parameters are announced. In what follows, when interpreting the results, we will use the estimates for the median bank. Table 2 provides our key parameter estimates for the two-stage structural game for whole Eurozone banks for time periods $2002-2007$ (first period) and $2008-2015$ (second period).[36] It turns out that Bayes factor favors the two-stage game against other two options (see Table 7). Hence, we concentrate on the two-stage structural model and only announce the estimates from structural models.[37] The columns show the estimation results for different percentiles of banks in terms of size, i.e., total assets. For example, 50% column gives the sample posterior mean of parameters for the 50^{th} percentile bank in terms of size, i.e., total assets. The median Eurozone bank's structural inefficiency in $2008-2015$ is 7.4%, which is much higher than that of $2002-2007$, i.e., 5.5%. Non-structural inefficiency, the inefficiency component that is not explained by managers actions, also follows similar trend and increases from 12.8% in $2002-2007$ to 16.2% in $2008-2015$ period. Also note that non-structural inefficiency is more than double of the structural inefficiency. For a median Eurozone bank, both structural and non-structural inefficiency increased significantly over time. Note that non-structural inefficiency, an inefficiency that has not been measured and thereby revealed by previous studies, is the unexplained inefficiency by our structural model. Such inefficiency is hard to tackle as little

[36] In all tables the γ_0 estimates are divided by $10,000$.

[37] The parameter estimates for the reduced form model are available upon request.

control can be applied to. It is of interest to further identify whether underlying heterogeneity in the Eurozone could drive such inefficiency in line with Joyce et al. (2012) who argue that banking imbalances in the Eurozone are rooted on the asymmetries between the periphery and the core of the Eurozone. Next, we turn our attention to the non-structural inefficiency at the periphery vs the core of the Eurozone.

The source of structural inefficiency is the game theoretical interactions between managers.

Table 2. Key Parameter Estimates for Two-Stage Structural Game: Eurozone

	2002-2007					2008-2015				
	10%	25%	50%	75%	90%	10%	25%	50%	75%	90%
α_1	0.312	0.332	0.324	0.205	0.144	0.212	0.202	0.225	0.217	0.315
	(0.015)	(0.021)	(0.017)	(0.020)	(0.014)	(0.014)	(0.012)	(0.017)	(0.019)	(0.020)
α_2	0.144	0.132	0.120	0.081	0.042	0.061	0.077	0.035	0.041	0.063
	(0.015)	(0.024)	(0.019)	(0.012)	(0.021)	(0.015)	(0.017)	(0.012)	(0.020)	(0.022)
β_1	0.014	0.022	0.031	0.028	0.017	0.31	0.028	0.033	0.077	0.021
	(0.001)	(0.005)	(0.017)	(0.002)	(0.003)	(0.007)	(0.005)	(0.007)	(0.012)	(0.005)
β_2	0.125	0.212	0.315	0.188	0.073	0.225	0.317	0.388	0.401	0.383
	(0.017)	(0.014)	(0.021)	(0.015)	(0.021)	(0.014)	(0.022)	(0.015)	(0.026)	(0.025)
γ_0	9.439	10.520	11.371	8.523	7.572	10.210	11.337	12.201	10.137	9.770
	(2.34)	(2.543)	(1.875)	(2.771)	(2.435)	(2.173)	(3.545)	(3.325)	(2.435)	(2.870)
γ_1	0.171	0.180	0.197	0.221	0.233	0.324	0.316	0.212	0.188	0.144
	(0.014)	(0.012)	(0.015)	(0.012)	(0.010)	(0.013)	(0.021)	(0.023)	(0.013)	(0.014)
SI	0.034	0.041	0.055	0.059	0.061	0.051	0.073	0.074	0.075	0.082
	(0.014)	(0.016)	(0.021)	(0.019)	(0.038)	(0.035)	(0.022)	(0.021)	(0.025)	(0.017)
NSI	0.182	0.155	0.128	0.120	0.155	0.212	0.197	0.162	0.164	0.180
	(0.021)	(0.024)	(0.019)	(0.027)	(0.020)	(0.025)	(0.022)	(0.017)	(0.011)	(0.017)

Note: We report sample means of posterior mean estimates. The percentage values in columns represent the size percentile of the relevant bank. Standard deviations are given in parenthesis. SI: Structural inefficiency and NSI: Non-structural inefficiency.

In Table 3, we present the key parameter estimates for the core vs the periphery of the Eurozone. The sample means of posterior mean estimates for structural and non-structural inefficiencies for both core and periphery countries increased over time. Interestingly, both inefficiencies increased in the aftermath of the financial crisis, whereas the non-structural inefficiency is more than twice that of the structural inefficiency for the both for the periphery and the core of the Eurozone. Our results show that the non-structural inefficiency is quite high for both the core and the periphery of the Eurozone at $18.7 - 18.8\%$ in the period $2008 - 2015$. This is of importance as it is for the first time that we confirm the discussion of Sinn and Wollmerchauser (2011) and Joyce et al. (2012) that predicted that 'target system imbalances', and as such outside the control of bank level management, would impair bank performance in the Eurozone. We estimate that non-structural inefficiency is indeed an issue with the Eurozone banking and after the crisis equally so for the core and the periphery. The core of the Eurozone may not be immune to non-structural inefficiency and as such any correction action would warrant that such imbalances should be challenged at a horizontal level. This action would imply that the way forward should be towards higher degree of integration that eventually could lead to a banking union. In fact, the reported evidence shows that the Eurozone as a currency union has done little to correct non-structural inefficiency, and in particular in the aftermath of the financial crisis, across all Member States.

We also announce the returns to scale and technical change estimates in Table 4. The returns to scale estimates are reasonable and in line with the estimates in the literature.[38]

[38] See Delis and Tsionas (2009) and Mamatzakis et al. (2015).

Table 3. Key Parameter Estimates for Two-Stage Structural Game: Core and Periphery

	2002-2007		2008-2015	
	CORE	PERIPHERY	CORE	PERIPHERY
α_1	0.313	0.177	0.144	0.132
	(0.014)	(0.013)	(0.021)	(0.014)
α_2	0.072	0.013	0.032	0.010
	(0.021)	(0.003)	(0.007)	(0.002)
β_1	0.035	0.010	0.015	0.011
	(0.008)	(0.003)	(0.004)	(0.002)
β_2	0.215	0.107	0.115	0.031
	(0.027)	(0.002)	(0.015)	(0.011)
γ_0	16.21	7.43	12.17	6.81
	(3.13)	(2.32)	(2.65)	(1.76)
γ_1	0.142	0.177	0.212	0.133
	(0.003)	(0.008)	(0.014)	(0.022)
SI	0.042	0.081	0.065	0.092
	(0.028)	(0.027)	(0.021)	(0.015)
NSI	0.162	0.175	0.188	0.187
	(0.017)	(0.013)	(0.018)	(0.020)

Note: We report sample means of posterior mean estimates. Standard deviations are given in parenthesis. SI: Structural inefficiency and NSI: Non-structural inefficiency.

Table 4. Returns to Scale and Technical Change Estimates

	Eurozone	Core	Periphery		
Returns to Scale	0.975	0.987	0.882		
	(0.023)	(0.021)	(0.017)		
Technical Change	-0.002	0.004	-0.001		
	(0.013)	(0.013)	(0.005)		
Eurozone	10%	25%	50%	75%	90%
Returns to Scale	0.889	0.912	0.925	0.984	0.999
	(0.011)	(0.013)	(0.017)	(0.018)	(0.021)
Technical Change	-0.005	-0.003	-0.001	0.002	0.005
	(0.007)	(0.013)	(0.015)	(0.010)	(0.015)

Note: We report sample means of posterior mean estimates. Standard deviations are given in parenthesis.

Table 5-6 show the simultaneous game estimation results for the key parameters. These estimates suggest increase in inefficiency of Eurozone banks over time as well. Also, note that Table 6 clearly shows that non-structural inefficiency is a serious impediment to bank performance for both core and periphery in the period 2008 to 2015 as it mounts at more than twice the structural inefficiency.

Table 5. Key Parameter Estimates for Simultaneous Structural Game: Eurozone

	2002-2007					2008-2015				
	10%	25%	50%	75%	90%	10%	25%	50%	75%	90%
α_1	0.185	0.144	0.221	0.232	0.255	0.141	0.183	0.120	0.081	0.077
	(0.022)	(0.017)	(0.023)	(0.015)	(0.017)	(0.021)	(0.018)	(0.015)	(0.014)	(0.014)
α_2	0.022	0.031	0.044	0.025	0.017	0.015	0.022	0.031	0.042	0.021
	(0.007)	(0.005)	(0.006)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.015)	(0.009)
β_1	0.017	0.025	0.031	0.044	0.032	0.018	0.022	0.025	0.028	0.031
	(0.013)	(0.015)	(0.014)	(0.012)	(0.015)	(0.004)	(0.008)	(0.003)	(0.015)	(0.005)
β_2	0.187	0.221	0.337	0.401	0.388	0.525	0.422	0.387	0.314	0.288
	(0.004)	(0.003)	(0.015)	(0.017)	(0.021)	(0.017)	(0.012)	(0.010)	(0.022)	(0.034)
γ_0	12.330	12.113	11.762	9.454	8.333	9.512	8.540	7.341	7.120	6.102
	(3.312)	(2.861)	(1.463)	(0.871)	(0.940)	(1.454)	(2.302)	(0.982)	(1.540)	(2.331)
γ_1	0.181	0.223	0.236	0.317	0.440	0.121	0.220	0.255	0.332	0.367
	(0.027)	(0.014)	(0.012)	(0.013)	(0.017)	(0.010)	(0.014)	(0.020)	(0.023)	(0.021)
SI	0.051	0.044	0.032	0.055	0.063	0.065	0.057	0.055	0.067	0.072
	(0.022)	(0.017)	(0.020)	(0.017)	(0.015)	(0.025)	(0.032)	(0.027)	(0.022)	(0.027)
NSI	0.151	0.172	0.185	0.193	0.196	0.185	0.194	0.222	0.204	0.214
	(0.023)	(0.019)	(0.028)	(0.021)	(0.020)	(0.023)	(0.020)	(0.018)	(0.014)	(0.019)

Note: We report sample means of posterior mean estimates. The percentage values in columns represent the size percentile of the relevant bank. Standard deviations are given in parenthesis. SI: Structural inefficiency and NSI: Non-structural inefficiency.

Table 6. Key Parameter Estimates for Simultaneous Structural Game: Core and Periphery

	2002-2007		2008-2015	
	CORE	PERIPHERY	CORE	PERIPHERY
α_1	0.315	0.132	0.132	0.117
	(0.022)	(0.014)	(0.004)	(0.005)
α_2	0.021	0.017	0.019	0.020
	(0.007)	(0.002)	(0.005)	(0.002)
β_1	0.013	0.012	0.011	0.015
	(0.004)	(0.006)	(0.006)	(0.004)
β_2	0.388	0.171	0.277	0.121
	(0.010)	(0.032)	(0.015)	(0.007)
γ_0	11.14	7.33	9.62	6.44
	(1.14)	(1.033)	(0.74)	(1.08)
γ_1	0.388	0.221	0.187	0.120
	(0.015)	(0.010)	(0.004)	(0.005)
SI	0.035	0.062	0.052	0.078
	(0.031)	(0.040)	(0.037)	(0.029)
NSI	0.171	0.193	0.195	0.213
	(0.024)	(0.020)	(0.024)	(0.020)

Note: We report sample means of posterior mean estimates. Standard deviations are given in parenthesis. SI: Structural inefficiency and NSI: Non-structural inefficiency.

In Table 7, we present Bayes factors in favor of two-stage structural game. A value larger than 1 indicates that the two-stage model is preferred. A widely used table for interpreting Bayes factors (Kass and Raftery, 1995) suggests that the values between 1 and 3 correspond to “Not worth more than a bare mention”; the values between 3 and 20 correspond to “Positive”; the values between 20 and 150 correspond to “Strong” evidence. Based on this table, for most of our model comparisons, the evidence for two-stage structural game is either positive or strong. The smallest Bayes factor in Table 7 is 4.72, which still presents positive evidence supporting the two-stage model. Hence, two-stage game is preferred over simultaneous game and reduced form model.

Table 7. Bayes Factors in Favour of Two-Stage Game

2002-2007	Eurozone	Core	Periphery
SG	4.72	8.15	14.32
RF	19.81	55.10	88.17
2008-2015	Eurozone	Core	Periphery
SG	17.32	92.44	121.55
RF	44.32	128.10	177.03

Note: SG: Simultaneous game; RF: Reduced form.

In Table 8, we present the rank correlation coefficients for the efficiency estimates from our models. As it appears, the efficiency estimates critically depend on the model choice. Thus, ignoring the two-stage structure of the decision mechanism may affect efficiency estimates in a way that affect rankings of banks. This seems to pose a serious threat to the standard stochastic frontier models. Hence, understanding and controlling for the structural mechanism that is determining the inefficiencies of banks appears to be vital when measuring inefficiency.

Table 8. Rank Correlation Coefficients Between Inefficiencies

	2002-2007					2008-2015				
With NSI	10%	25%	50%	75%	90%	10%	25%	50%	75%	90%
SG & TSG	0.112	0.107	0.102	0.113	0.115	0.131	0.032	0.015	0.014	0.060
SG & RF	0.004	0.015	0.013	0.019	0.020	0.018	0.020	0.012	0.016	0.013
TSG & RF	0.011	0.012	0.010	0.011	0.012	0.010	0.005	0.001	0.002	0.003
Without NSI	10%	25%	50%	75%	90%	10%	25%	50%	75%	90%
SG & TSG	0.121	0.114	0.120	0.115	0.007	0.006	0.087	0.013	0.017	0.012
SG & RF	0.015	0.007	0.012	0.015	0.007	0.015	0.001	0.001	0.002	0.003
TSG & RF	0.014	0.011	0.015	0.017	0.005	0.001	0.002	0.002	0.002	0.002

Note: SG: Simultaneous game; TSG: Two-stage game; RF: Reduced form, SI: Structural inefficiency and NSI: Non-structural inefficiency.

4 Concluding Remarks

As argued by Evanoff and Israilevich (1991), it is possible to identify a number of possible explanations for the inefficiency in banking industry. Most of these reasons can be found in any industry. Some of the earlier studies argue that bank size would be a determinant of inefficiency. Although larger banks likely to have more complex structure, they also would have higher pressure from owners. Also, these banks are in more competitive markets, which may result in less inefficiency production process.

Frantz (2014) explains Leibenstein's view for why inefficiency may be present. First, the human personality has a superego and an id. If we follow the former, we would work hard and try to do things "correctly." On the other hand, if we follow the latter, we would avoid anything requiring focused attention, which relates to lazy decision rules. Second, in general, managers are not the owners, which leads to an agency problem. Third, the workers can choose their effort levels. Fourth, monopoly power triggers these factors, especially the second and third ones.

Our study relies on the literature that considers competition and efficiency simultaneously and proposes a new approach whereby this is examined through a novel principal agent model. Unlike our model, most principal agent studies do not incorporate agent based competition. Hence, our theoretical model combines competition, principal agent problem, and efficiency. We decompose and estimate structural and nonstructural efficiency for the first time to the best of our knowledge. The importance of such a decomposition is that we offer an understanding to what extent efficiency can be controlled through policies that affect competition or managerial compensation.

Therefore, our paper provided agency based models for understanding the sources of inefficiency and, in particular bank inefficiency, further. We identified several main factors: Market power, unobserved characteristics of managers, terms of contract, heterogeneity across Member States of the Eurozone, in particular between the periphery and the core, and the "average inefficiency." In the standard stochastic frontier models, the first factor can be controlled by modeling the distribution of the inefficiency as a function of number of banks or another measure of market power. However, the systematic nature of the remaining factors makes the identification of inefficiency difficult in the conventional reduced form stochastic frontier models. We illustrated that parameter estimates for the conventional stochastic cost frontier models would not be consistent due to misspecification. This poses a serious problem for empirical applications that estimate a cost function.

We applied our novel econometric structural and reduced form models to the Eurozone banks between 2002 and 2015. The inefficiency of these banks are important for at least two reasons. First, if the banks perform efficiently, they will be more successful in surviving. Wheelock and Wilson (2000) find that those banks with higher technical inefficiency, have higher likelihood of failure. Second, bank efficiency analysis enables us to understand the effects of financial integration such as in a currency union as the Eurozone, which could help forming better strategies for picking the pace of further integration.

Bayes factor favors the sequential model where the inefficiency is determined first. This means that the implementation of proper monitoring mechanisms and of managerial contracts plays an essential role in the inefficiency levels. In addition, moving towards a banking union would enhance the integration and the homogeneity across all Member States in the Eurozone and would assist efforts to subdue the non-structural inefficiency. Since inefficiency is a form of social welfare loss, these corrective policy actions would not only help the banks to improve their profits but also decrease waste of resources through efficiency gains.

References

Aigner, D.J., Lovell, C.A.K., and Schmidt, P. (1977), Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 21-37.

Allen, L. and Rai, A. (1996), Operational Efficiency in Banking: An International Comparison, *Journal of Banking and Finance*, 20, 655-672.

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, 69-84.

Altunbas, Y., Gardener, E.P.M., Molyneux, P., and Moore, B. (2001), Efficiency in European banking, *European Economic Review,* 45, 1931-1955.

Amir, R. (1996), Cournot Oligopoly and the Theory of Supermodular Games, *Games and Economic Behavior*, 15:132-148.

Andrieş, A.M. and Căpraru, B. (2014), The Nexus Between Competition and Efficiency: The European Banking Industries Experience, *International Business Review*, 23, 566-579.

Assaf, A.G., Bu, R., and Tsionas, M.G. (2019), A Bayesian Approach to Continuous Type Principal-Agent Problems, Forthcoming in *European Journal of Operational Research.*

Baek, H.Y. and Pagán, J.A. (2002), Executive Compensation and Corporate Production Efficiency: A Stochastic Frontier Approach, Quarterly Journal of Business and Economics, 41, 27-41.

Berger, A.N. and DeYoung, R. (1997), Problem Loans and Cost Efficiency in Commercial Banking, *Journal of Banking and Finance*, 21, 6, 849-870.

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, 454-465.

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, 175-212.

Berr, F. (2011), Stackelberg Equilibria in Managerial Delegation Games, *European Journal of Operational Research*, 212, 251-262.

Bikker, J.A. and Bos, J.W. (2008), Bank Performance: A Theoretical and Empirical Framework for the Analysis of Profitability, Competition, and Efficiency, Routledge International Studies in Money and Banking.

Brissimis, S.N., Delis, M.D, and Tsionas, E.G. (2010), Technical and Allocative

Efficiency in European Banking, *European Journal of Operational Research,* 204, 153-163.

Casella, G. and George, E.I. (1992), Explaining the Gibbs Sampler, *The American Statistician*, 46, 167-174.

Cecchini, M., Ecker, J., Kupferschmid, M., Leitch, R. (2013), Solving Nonlinear Principal-Agent Problems Using Bilevel Programming, *European Journal of Operational Research*, 230, 364-373.

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, 185-200.

Corts, K.S. (1999), Conduct Parameters and the Measurement of Market Power, *Journal of Econometrics*, 88, 227-250.

Delis, M., Iosifidi, M., and Tsionas, M.G. (2017), Endogenous Bank Risk and Efficiency, *European Journal of Operational Research*, 260, 376-387.

Delis, M.D. and Tsionas, E.G. (2009), The Joint Estimation of Bank-Level Market Power and Efficiency, *Journal of Banking and Finance*, 33, 1842-1850.

Demstez, H. and Lehn, K. (1985), The Structure of Corporate ownership: Causes and Consequences, *Journal of Political Economy*, 93, 1155-1177.

Dong, Y., Firth, M., Hou, W., and Yang, W. (2016), Evaluating the performance of Chinese commercial banks: A comparative analysis of different types of banks, *European Journal of Operational Research*, 252, 280-295.

Duygun, M., Kutlu, L., and Sickles, R.C. (2016) Measuring Productivity and Efficiency: A Kalman Filter Approach, *Journal of Productivity Analysis*, 46, 155-167.

Evanoff, D.D. and Israilevich, P.R. (1991), Productive Efficiency in Banking, *Economic Perspectives*, Federal Reserve Bank of Chicago, 11-32, July/August.

Farrell, J. and Shapiro, C. (1990), Horizontal Mergers: An Equilibrium Analysis, *American Economic Review*, 80, 107-126.

Festinger, L. (1954), A Theory of Social Comparison Processes, *Human Relations*, 7, 117-140.

Francis, B., Gupta, A, Hasan, I. (2015), Impact of Compensation Structure and Managerial Incentives on Bank Risk Taking, *European Journal of Operational Research*, 242, 651-676.

Frantz, R. (2014), X-(In)efficiency in Banking 1995-2011, Paper Prepared for Annual Conference of the American Antitrust Institute, Washington, D.C., Conference Theme: "Efficiency."

Fuentes, R. and Vergara, M. (2007), Is Ownership Structure a Determinant of Bank Efficiency?, Central Bank of Chile Working Papers, No 456.

Fujii, H., Managi, S., Matousek, R., and Rughoo, A. (2017), Bank Efficiency, Productivity and Convergence in Europe, MPRA Working Paper, 77237.

Gagnepain, P. and Ivaldi, M. (2002), Incentive Regulatory Policies: The Case of Public Transit Systems in France, *RAND Journal of Economics*, 33, 605-29.

Galan, G.E., Veiga, H., and Wiper, M.P. (2015), Dynamic effects in inefficiency: Evidence from the Colombian banking sector, *European Journal of Operational Research*, 250, 562-571.

Gathon, H.-J. and Pestieau, P. (1995), Decomposing Efficiency Into its Managerial and its Regulatory Components: The Case of European Railways, *European Journal of Operational Research*, 80, 500-507.

Gaudet, G. and Salant, S.W. (1991), Increasing the Profits of a Subset of Firms in Oligopoly Models with Strategic Substitutes, *American Economic Review*, 81, 658-665.

Gelfand, A.E. and Smith, A.F.M. (1990), Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, 85, 398-409.

Girolami, M. and Calderhead, B. (2011), Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods, *Journal of the Royal Statistical Society: Series B*, 73, 123-214.

Glass, A.J. and Kenjegalieva, K. (2019), A Spatial Productivity Index in the Presence of Efficiency Spillovers: Evidence for U.S. banks, 1992–2015, *European Journal of Operational Research*, 273, Issue 3, 1165-1179.

Green, A. and Mayes, D. (1991), Technical Inefficiency in Manufacturing Industries, *Economic Journal*, 101, 523-538.

Greene, W.H. (1980a), Maximum Likelihood Estimation of Econometric Frontier Functions, *Journal of Econometrics*, 3, 27-56.

Greene, W.H. (1980b), On the Estimation of a Flexible Frontier Production Model, *Journal of Econometrics*, 13, 101-115.

Greene, W.H. (2003), Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function, *Journal of Productivity Analysis*, 19, 179-190.

Hanousek, J., Kočenda, E., and Shamshur, A. (2015), Corporate Efficiency in Europe, Journal of Corporate Finance, 32, 24-40.

Hicks, J.R. (1935), Annual Survey of Economic Theory: The Theory of Monopoly, *Econometrica*, 3, 1-20.

Hoffmann, A. and Post, T. (2013), How Return and Risk Experiences Shape Investors Belief and Preferences, Netspar Discussion Paper, No. 11/2012-044.

Hudomiet, P., Kézdi, G., and Willis, R. (2011), Stock Market Crash and Expectations of American Households, *Journal of Applied Econometrics*, 26, 393-415.

Joyce, M., Miles, D., Scott, A., and Vayanos, D. (2012), Quantitative Easing and Unconventional Monetary Policy – an Introduction, *Economic Journal*, 122, F271-F288.

Kauko, K. (2009), Managers and Efficiency in Banking, *Journal of Banking and Finance*, 33, 546-556.

Kass, R. and Raftery, A. (1995), Bayes Factors, *Journal of American Statistical Association*, 90, 773-795.

Koetter, M., Kolari, J.W., and Spierdijk, L. (2012), Enjoying the Quiet Life Under Deregulation? Evidence from Adjusted Lerner Indices for U.S. Banks, *Review of Economics and Statistics*, 94, 462-480.

Krishna, P. (2001), On Competition and Endogenous firm Efficiency, *Economic Theory*, 18, 753-60.

Kutlu, L. (2012), US Banking Efficiency, 1984-1995, *Economics Letters*, 117, 53-56.

Kutlu, L. (2017), A Constrained State Space Approach for Estimating Firm Efficiency, *Economics Letters*, 152, 54-56.

Kutlu, L. and McCarthy, P. (2016), US Airport Governance and Efficiency, *Transportation Research Part E*, 89, 117-132.

Kutlu, L. and Nair-Reichert, U. (2020), Efficiency and Executive Pay: Evidence from Indian Manufacturing Sector, Working Paper.

Kutlu, L. and Sickles, C.R. (2012), Estimation of Market Power in the Presence of Firm Level Inefficiencies, *Journal of Econometrics*, 168, 141-155.

Kutlu, L. and Wang, R. (2018a), Estimation of Cost Efficiency Without Cost Data, *Journal of Productivity Analysis*, 49, 137-151.

Kutlu, L. and Wang, R. (2018b), Price Dispersion, Competition, and Efficiency - Evidence from Chicago Based Routes, *Journal of Transport, Economics and Policy*, 52, 23-44.

Levinsohn, J. and Petrin, A. (2003), Estimating Production Functions Using Inputs to Control for Unobservables, *Review of Economic Studies*, 70, 317-341.

Liang, Q., Xu, P., and Jiraporn, P. (2013), Board Characteristics and Chinese Bank Performance, *Journal of Banking and Finance*, 37, 2953-2968.

Lozano-Vivas, .A. (1997), Profit efficiency for Spanish savings banks, *European Journal of Operational Research*, 98, 381-394.

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, 59-77.

Mamatzakis, E., Tsionas, M.G., Kumbhakar, S.C., and Koutsomanoli-Filippaki, A. (2015), Does Labour Regulation Affect Equally Technical and Allocative Efficiency? Evidence From the Banking Industry, *Journal of Banking and Finance*, 61, S84-S98.

Martin, S. (1993), Endogenous Firm Efficiency in a Cournot Principal-Agent Model, *Journal of Economic Theory*, 59, 445-450.

Maudos, J. and de Guevara, J.F. (2007), The Cost of Market Power in Banking: Social Welfare Loss vs. Cost Inefficiency, *Journal of Banking and Finance*, 31, 2103-2135.

Maudos, J., Pastor, J.M., Peréz, F., and Quesada, J. (2002), Cost and Profit Efficiency in European Banks, *Journal of International Financial Markets, Institutions and Money,* 12, 33-58.

Meeusen, W. and van den Broeck, J. (1977), Efficiency Estimation from Cobb–

Douglas Production Functions with Composed Error, *International Economic Review*, 18, 435-444.

Mester, L.J. (1993), Efficiency in the Savings and Loan Industry, *Journal of Banking and Finance*, 17, 267-286.

Muñiz, M.A. (2002), Separating Managerial Inefficiency and External Conditions in Data Envelopment Analysis, *European Journal of Operational Research*, 143, 625-643.

Norde, H., Özen, U., and Slikker, M. (2016), Setting the Right Incentives for Global Planning and Operations, *European Journal of Operational Research*, 253, 441-455.

Novshek, W. (1985), On the Existence of Cournot Equilibrium, *Review of Economic Studies*, 52, 85-98.

Olley, S. and Pakes, A. (1996), The Dynamics of Productivity in the Telecommunications Equipment Industry, *Econometrica,* 64,1263-1295.

Paradi, J.C., Zhu, H., and Edelstein, B. (2012), Identifying Managerial Groups in a Large Canadian Bank Branch Network with a DEA Approach, *European Journal of Operational Research*, 219, 178-187.

Pi, L. and Timme, S.G. (1993), Corporate Control and Bank Efficiency, *Journal of Banking and Finance*, 17, 515-530.

Seade, J.K. (1980), The Stability of Cournot Revisited, *Journal of Economic Theory*, 23, 15-27.

Sealey, C.W., Jr., and Lindley, J.T. (1977), Inputs, Outputs, and Theory of Production Cost at Depository Financial Institutions, *Journal of Finance*, 32, 1251-1266.

Schmidt, P., and Sickles, R.C. (1984), Production Frontiers and Panel Data, *Journal of Business and Economic Statistics*, 2, 367-374.

Sealey, C. and Lindley, J. (1977), Inputs, Outputs and a Theory of Production and Cost of Depository Financial Institutions, *Journal of Finance*, 32, 1251-266.

Shapiro, C. (1989), Theories of Oligopoly Behavior, in Handbook of Industrial Organization, vol. I, chapter 6. North Holland, Amsterdam.

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

Sinn, H.-W. and Wollmershauser, T. (2011)., Target Loans, Current Account Balances and Capital Flows: The ECB's Rescue Facility', CESIFO Working Paper No. 3500.

Smith, A.F.M. and Roberts, C.O. (1993), Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of Royal Statistical Society* B55, 2-23.

Stevenson, R. (1980), Likelihood Functions for Generalized Stochastic Frontier Functions, *Journal of Econometrics*, 13, 57-66.

Tanner, M.A. and Wong, W.H. (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, 82, 528-540.

Tsionas, E.G. (2002), Stochastic Frontier Models with Random Coefficients, *Journal of Applied Econometrics*, 17, 127-147.

Tsionas, M.G. (2017), Microfoundations for Stochastic Frontiers, *European Journal of Operational Research*, 258, 1165-1170.

Tsionas, E.G., Malikov, E., and Kumbhakar, S.C. (2018), An Internally Consistent Approach to the Estimation of Market Power and Cost Efficiency with an Application to U.S. Banking, *European Journal of Operational Research*, 270, 747-760.

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, 129-144.

Wheelock, D.C. and Wilson, P.W. (2000), Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions, *Review of Economics and Statistics*, 82, 127-138.

Williams, J. (2004), Determining Management Behavior on European banking, *Journal of Banking and Finance*, 28, 2427-2460.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics, Wiley, New York.

5 Appendix A: Propositions

Proposition 1 presents the equilibrium conditions and profits of firms conditional on $\mathbf{r}$.[39] In Propositions 2-3, we analyze the responses of market quantity and profit of a firm to a change in r_i.[40] These propositions would help us understand the characteristics of the two-stage game. Note that in Proposition 1-3, we assume that the marginal cost is constant, i.e., $C_i(q_i) = cq_i$. This assumption is made to get closed form solutions for equilibrium. Most of our analysis does not require

[39] Conditional on $\mathbf{r}$, the equations given in Proposition 1 are standard Cournot solutions. In particular, Equation (13) follows from Equation (2) of Farrell and Shapiro (1990); and the Cournot-Nash quantity and profit of firm i, directly follows from this equation. Hence, we skip the proof.

[40] We provide the proofs of Proposition 2-3 in Appendix A, which is available upon request.

this assumption.[41] For example, in our empirical model, we assume a translog functional form for the cost function.

Proposition 1: Assume that the marginal cost is constant so that $C_i(q_i) = cq_i$ and $\theta(\mathbf{r}) \equiv \frac{Q(r)P''(Q(\mathbf{r}))}{P'(Q(\mathbf{r}))} \geq -2$.[42] The Nash equilibrium market output, $Q(\mathbf{r})$, conditional on $\mathbf{r}$ is the solution of:

$$nP(Q(\mathbf{r})) + Q(\mathbf{r})P'(Q(\mathbf{r})) = nc\bar{r}. \tag{13}$$

The Cournot-Nash quantity of firm i is given by:

$$q_i(\mathbf{r}) = \frac{c(\bar{r} - r_i)}{-P'(Q(\mathbf{r}))} + \frac{Q(\mathbf{r})}{n} \tag{14}$$

The Cournot-Nash profit of firm i is given by:

$$\begin{aligned} \pi_i(\mathbf{q}, \mathbf{r}) &= -P'(Q(\mathbf{r}))\, q_i^2(\mathbf{r}) \\ &= -P'(Q(\mathbf{r})) \left(\frac{c(\bar{r} - r_i)}{-P'(Q(\mathbf{r}))} + \frac{Q(\mathbf{r})}{n} \right)^2. \end{aligned} \tag{15}$$

Proposition 2: Assume that the marginal cost is constant so that $C_i(q_i) = cq_i$ and $\theta(\mathbf{r}) \equiv \frac{Q(r)P''(Q(\mathbf{r}))}{P'(Q(\mathbf{r}))} \geq -2$. In the second stage, the market output decreases

[41] Whenever this assumption is needed, we explicitly mention as in Proposition 1-3.

[42] This assumption (i.e., $C_i(q_i) = cq_i$ and $\theta(\mathbf{r}) \equiv \frac{Q(r)P''(Q(\mathbf{r}))}{P'(Q(\mathbf{r}))} \geq -2$) assures that the Nash equilibrium in the second stage exists and is unique. For example, for linear demand, we have $\theta(\mathbf{r}) = 0$; and for constant elasticity demand $P(Q(\mathbf{r})) = \alpha Q(\mathbf{r})^{-\beta}$, we have $\theta(\mathbf{r}) = -(1+\beta)$. It is possible to relax the constant marginal cost assumption if we are ready to restrict $\theta \geq -1$. For example, Farrell and Shapiro (1990) and Gaudet and Salant (1991) assumed $\theta > -1$. See also Seade (1980), Novshek (1985), Shapiro (1989), and Amir (1996) for more general conditions for existence.

as lack of effort, r_i, increases:

$$\frac{\partial Q(\mathbf{r})}{\partial r_i} = \frac{c}{P'(Q(\mathbf{r}))(n+1+\theta(\mathbf{r}))} < 0. \tag{16}$$

In particular, firm level quantities satisfy:

$$\begin{aligned} \frac{\partial q_i(\mathbf{r})}{\partial r_i} &= \frac{(n+\theta(\mathbf{r})(1-s_i(\mathbf{r})))c}{P'(Q(\mathbf{r}))(n+1+\theta(\mathbf{r}))} < 0 \\ \frac{\partial q_i(\mathbf{r})}{\partial r_j} &= -\frac{(1+\theta(\mathbf{r})s_i(\mathbf{r}))c}{P'(Q(\mathbf{r}))(n+1+\theta(\mathbf{r}))} \begin{array}{ll} \leq 0 & \text{if} \quad s_i(\mathbf{r}) \geq 1/2 \\ > 0 & \text{if} \quad s_i(\mathbf{r}) < 1/2 \end{array}, \end{aligned} \tag{17}$$

where $s_i(\mathbf{r}) = \frac{q_i(\mathbf{r})}{Q(\mathbf{r})}$ is the market share of firm i.

In Proposition 2, we observe that q_i and Q decrease as the lack of effort, r_i, increases. The sign of $\frac{\partial q_i(\mathbf{r})}{\partial r_j}$ is the same as the sign of $(1+\theta(\mathbf{r})s_i(\mathbf{r}))$, which is ambiguous. However, $\frac{\partial q_i(\mathbf{r})}{\partial r_j} > 0$ if $s_i(\mathbf{r}) < 1/2$. Therefore, in most cases, as the other managers get more inefficient, q_i increases.

Then, the first order conditions for the first stage are:

$$\begin{aligned} \frac{\partial U_i(q(\mathbf{r}), r_i)}{\partial r_i} &= \frac{\partial \pi_i(\mathbf{q}, \mathbf{r})}{\partial r_i} + V_i'(r_i) = 0 \Rightarrow \\ V_i'(r_i) &= -\frac{\partial (P(Q(\mathbf{r}))q_i(\mathbf{r}) - C_i(q_i(\mathbf{r}))r_i)}{\partial r_i}. \end{aligned} \tag{18}$$

In this model, we have r_i appearing in the system and this is somewhat difficult to solve explicitly. We can still use the equation:

$$P(Q(\mathbf{r})) = -P'(Q(\mathbf{r}))q_i(\mathbf{r}) + C_i'(q_i)r_i \tag{19}$$

and obtain a solution for r_i using a standard line search procedure. From this equation the quantities can be solved numerically as a function of $\mathbf{r}$. Then, the first stage first order conditions in Equation (18) can be used to solve for $\mathbf{r}$ numerically.

Proof of Proposition 2: The first order condition for the second stage is:

$$\frac{\partial \pi_i(\mathbf{q}, \mathbf{r})}{\partial q_i} = P'(Q(\mathbf{r}))\, q_i(\mathbf{r}) + P(Q(\mathbf{r})) - c r_i = 0. \tag{20}$$

After summing Equation (20) over i, we get:

$$\begin{aligned}
nP(Q(\mathbf{r})) + Q(\mathbf{r})\, P'(Q(\mathbf{r})) &= nc\bar{r} \Rightarrow \\
\frac{\partial Q(\mathbf{r})}{\partial r_i} P'(Q(\mathbf{r})) \left(n + 1 + \frac{Q(\mathbf{r})\, P''(Q(\mathbf{r}))}{P'(Q(\mathbf{r}))} \right) &= c \Rightarrow \\
\frac{\partial Q(\mathbf{r})}{\partial r_i} &= \frac{c}{P'(Q(\mathbf{r}))\,(n + 1 + \theta(\mathbf{r}))} < 0.
\end{aligned}$$

In order to get $\frac{\partial q_i(\mathbf{r})}{\partial r_i}$, we differentiate Equation (20) with respect to r_i:

$$\begin{aligned}
\frac{\partial^2 \pi_i(\mathbf{q}, \mathbf{r})}{\partial q_i \partial r_i} &= 0 \Rightarrow \\
P''(Q(\mathbf{r})) \frac{\partial Q(\mathbf{r})}{\partial r_i} q_i(\mathbf{r}) + P'(Q(\mathbf{r})) \frac{\partial q_i(\mathbf{r})}{\partial r_i} + P'(Q(\mathbf{r})) \frac{\partial Q(\mathbf{r})}{\partial r_i} &= c \Rightarrow \\
\left(P''(Q(\mathbf{r}))\, q_i(\mathbf{r}) + P'(Q(\mathbf{r})) \right) \frac{\partial Q(\mathbf{r})}{\partial r_i} + P'(Q(\mathbf{r})) \frac{\partial q_i(\mathbf{r})}{\partial r_i} &= c \Rightarrow \\
\frac{\partial q_i(\mathbf{r})}{\partial r_i} &= \frac{n + \theta(\mathbf{r})\,(1 - s_i(\mathbf{r}))}{P'(Q(\mathbf{r}))\,(n + 1 + \theta(\mathbf{r}))} c < 0.
\end{aligned}$$

In order to get $\frac{\partial q_i(\mathbf{r})}{\partial r_j}$, we differentiate Equation (20) with respect to r_j:

$$\begin{aligned}
\frac{\partial^2 \pi_i (\mathbf{q}, \mathbf{r})}{\partial q_i \partial r_j} &= 0 \Rightarrow \\
\left(P'' (Q (\mathbf{r})) q_i (\mathbf{r}) + P' (Q (\mathbf{r}))\right) \frac{\partial Q (\mathbf{r})}{\partial r_j} + P' (Q (\mathbf{r})) \frac{\partial q_i (\mathbf{r})}{\partial r_j} &= 0 \Rightarrow \\
\frac{\partial q_i (\mathbf{r})}{\partial r_j} &= - (1 + \theta (\mathbf{r}) s_i (\mathbf{r})) \frac{\partial Q (\mathbf{r})}{\partial r_j} \Rightarrow \\
\frac{\partial q_i (\mathbf{r})}{\partial r_j} &= \frac{- (1 + \theta (\mathbf{r}) s_i (\mathbf{r})) c}{P' (Q (\mathbf{r})) (n + 1 + \theta (\mathbf{r}))}.
\end{aligned}$$

Proposition 3: Assume that the marginal cost is constant so that $C_i (q_i) = cq_i$ and $\theta (\mathbf{r}) \equiv \frac{Q(r)P''(Q(\mathbf{r}))}{P'(Q(\mathbf{r}))} \geq -2$. In the second stage, the profit decreases as lack of effort, r_i, increases:

$$\frac{\partial \pi_i (\mathbf{q}, \mathbf{r})}{\partial r_i} = -q_i (\mathbf{r}) c \frac{2n + \theta (\mathbf{r}) (2 - s_i (\mathbf{r}))}{n + 1 + \theta (\mathbf{r})} < 0. \tag{21}$$

Although, higher inefficiency increases V_i, the profit of firm decreases by higher inefficiency levels. Hence, the optimal inefficiency level of a manager is determined by these conflicting effects. Now, we consider the optimization problem of managers in the first stage where they choose the optimal level for lack of effort:

$$\max_{r_i} U_i (q (\mathbf{r}), \mathbf{r}). \tag{22}$$

Then, the first order conditions for the first stage of two-stage game are:

$$\begin{aligned}\frac{\partial U_i\left(q\left(\mathbf{r}\right), r_i\right)}{\partial r_i} &= \frac{\partial \pi_i\left(\mathbf{q}, \mathbf{r}\right)}{\partial r_i} + V_i'\left(r_i\right) = 0 \Rightarrow \\ V_i'\left(r_i\right) &= -\frac{\partial\left(P\left(Q\left(\mathbf{r}\right)\right) q_i\left(\mathbf{r}\right) - C_i\left(q_i\left(\mathbf{r}\right)\right) r_i\right)}{\partial r_i}.\end{aligned}$$

The first order conditions of simultaneous game for q_i remains the same as the two-stage game and they are given by:

$$\begin{aligned}\frac{\partial U_i\left(\mathbf{q}, \mathbf{r}\right)}{\partial q_i} &= P'\left(Q\left(\mathbf{r}\right)\right) q_i\left(\mathbf{r}\right) + P\left(Q\left(\mathbf{r}\right)\right) - C_i'\left(q_i\left(\mathbf{r}\right)\right) r_i = 0 \qquad (23) \\ \frac{\partial U_i\left(\mathbf{q}, \mathbf{r}\right)}{\partial r_i} &= -C_i(q_i) + V_i'\left(r_i\left(q_i\right)\right) = 0.\end{aligned}$$

Proof of Proposition 3: In order prove Proposition 3, we find the derivative of (second-stage) profit with respect to inefficiency:

$$\begin{aligned}\frac{\partial \pi_i\left(\mathbf{q}, \mathbf{r}\right)}{\partial r_i} &= -\frac{\partial P'\left(Q\left(\mathbf{r}\right)\right) q_i^2\left(\mathbf{r}\right)}{\partial r_i} \\ &= -P''\left(Q\left(\mathbf{r}\right)\right) \frac{\partial Q\left(\mathbf{r}\right)}{\partial r_i} q_i^2\left(\mathbf{r}\right) - 2 q_i\left(\mathbf{r}\right) P'\left(Q\left(\mathbf{r}\right)\right) \frac{\partial q_i\left(\mathbf{r}\right)}{\partial r_i} \\ &= -q_i\left(\mathbf{r}\right) c \left(\frac{2n + 2\theta\left(\mathbf{r}\right) - \theta\left(\mathbf{r}\right) s_i\left(\mathbf{r}\right)}{n + 1 + \theta\left(\mathbf{r}\right)}\right) \\ &= -\theta\left(\mathbf{r}\right) \frac{c}{n + 1 + \theta\left(\mathbf{r}\right)} s_i\left(\mathbf{r}\right) q_i\left(\mathbf{r}\right) - 2 q_i\left(\mathbf{r}\right) \frac{n + \theta\left(\mathbf{r}\right)\left(1 - s_i\left(\mathbf{r}\right)\right)}{n + 1 + \theta\left(\mathbf{r}\right)} c \\ &= -q_i\left(\mathbf{r}\right) c \frac{2n + \theta\left(\mathbf{r}\right)\left(2 - s_i\left(\mathbf{r}\right)\right)}{n + 1 + \theta\left(\mathbf{r}\right)} < 0.\end{aligned}$$

6 Appendix B: Multioutput Case

While a single output model is useful in many situations, for some contexts, such as banking industry, a single output model may not be realistic. Hence, we extend our models to the multi-output setting. Without loss of generality, we will present a two-output scenario. We assume that there are two outputs with corresponding demands represented by $P_1(Q)$ and $P_2(Q)$ where $Q = (Q_1, Q_2)$ is the vector of total quantities for the products. For the two-stage structural game, the solution of the two-output case is similar. The relevant first order conditions for optimization problem of the managers in the second stage are given by:

$$\begin{aligned} P_1(Q(\mathbf{r})) &= -\frac{\partial P_1(Q(\mathbf{r}))}{\partial Q_1} q_{1i}(\mathbf{r}) - \frac{\partial P_2(Q(\mathbf{r}))}{\partial Q_1} q_{2i}(\mathbf{r}) + \frac{\partial C_i(q_i(\mathbf{r}))}{\partial q_{1i}} r_i \\ P_2(Q(\mathbf{r})) &= -\frac{\partial P_1(Q(r))}{\partial Q_2} q_{1i}(\mathbf{r}) - \frac{\partial P_2(Q(r))}{\partial Q_2} q_{2i}(\mathbf{r}) + \frac{\partial C_i(q_i(\mathbf{r}))}{\partial q_{2i}} r_i. \end{aligned} \tag{24}$$

Again, from these equations, the quantities can be solved numerically as a function of $\mathbf{r}$. Then, the first stage first order conditions:

$$V_i'(r_i) = -\frac{\partial\left(P_1(Q(\mathbf{r}))\, q_{1i}(\mathbf{r}) + P_2(Q(\mathbf{r}))\, q_{2i}(\mathbf{r}) - C_i(q_i(\mathbf{r}))\, r_i\right)}{\partial r_i} \tag{25}$$

can be used to solve for $\mathbf{r}$ numerically.

For the simultaneous structural game, the relevant first order conditions (with added error terms) are given by:

$$\begin{aligned}
P_{1i} &= -\frac{\partial P_{1i}}{\partial Q_{1i}} q_{1i} - \frac{\partial P_{2i}}{\partial Q_{1i}} q_{2i} + \frac{\partial C_i}{\partial q_{1i}} V_i'^{-1}(C_i) + v_{1i} \\
P_{2i} &= -\frac{\partial P_{1i}}{\partial Q_{2i}} q_{1i} - \frac{\partial P_{2i}}{\partial Q_{2i}} q_{2i} + \frac{\partial C_i}{\partial q_{2i}} V_i'^{-1}(C_i) + v_{2i} \\
\ln C_i^* &= \ln C_i + \ln V_i'^{-1}(C_i) + u_i + v_{3i},
\end{aligned} \tag{26}$$

where v_{1i} and v_{2i} are two-sided error terms, $u_i \geq 0$ is the one-sided error term that is capturing the non-structural inefficiency, and $\ln V_i'^{-1}(C_i)$ is the structural term that is capturing the structural inefficiency. These expressions in System (26) can be simultaneously estimated along with the demand equations $P_1(Q; X_d)$ and $P_2(Q, X_d)$ where X_d are exogenous demand shifters.

Appendix C: Structural Model Estimation

We use a Girolami and Calderhead (2011) (GC) algorithm to update draws for a parameter $\vartheta \in R^d$. The algorithm uses local information about both the gradient and the Hessian of the log-posterior conditional of ϑ at the existing draw. A Metropolis test is again used for accepting the candidate so generated but the GC algorithm moves considerably faster relative to our naive scheme previously described. The GC algorithm is started at the first-stage GMM estimator and MCMC is run until convergence. It has been found that the GC algorithm performs vastly superior relative to the standard MH algorithm and autocorrelations are much smaller. Suppose we have a parameter vector $\vartheta \in R^d$ and data $\boldsymbol{X}$.

Suppose $L(\vartheta) = \ln p(\vartheta \mid \boldsymbol{X})$ is used to denote for simplicity the log posterior of ϑ. Moreover, define:

$$\boldsymbol{G}(\vartheta) = \text{est.cov}\frac{\partial}{\partial\vartheta}\ln p\left(\boldsymbol{X} \mid \vartheta\right) \tag{27}$$

the empirical counterpart of

$$\boldsymbol{G}_o(\vartheta) = -E_{Y|\vartheta}\frac{\partial^2}{\partial\vartheta\partial\vartheta'}\ln p\left(\boldsymbol{X} \mid \vartheta\right). \tag{28}$$

The Langevin diffusion is given by the following stochastic differential equation:

$$d\vartheta\left(t\right) = \tfrac{1}{2}\tilde{\nabla}_{\vartheta}L\left\{\vartheta\left(t\right)\right\}dt + d\mathbf{B}\left(t\right), \tag{29}$$

where

$$\tilde{\nabla}_{\vartheta}L\left\{\vartheta\left(t\right)\right\} = -\boldsymbol{G}^{-1}\left\{\vartheta\left(t\right)\right\}\cdot\bigtriangledown_{\vartheta}L\left\{\vartheta\left(t\right)\right\} \tag{30}$$

is the so called "natural gradient" of the Riemann manifold generated by the log-posterior. The elements of the Brownian motion are

$$\begin{aligned}\boldsymbol{G}^{-1}\left\{\vartheta\left(t\right)\right\}d\mathbf{B}_i\left(t\right) =& \left|\boldsymbol{G}\left\{\vartheta\left(t\right)\right\}\right|^{-1/2}\sum_{j=1}^{K_\vartheta}\frac{\partial}{\partial\vartheta}\left[\boldsymbol{G}^{-1}\left\{\vartheta\left(t\right)\right\}_{ij}\left|\boldsymbol{G}\left\{\vartheta\left(t\right)\right\}\right|^{1/2}\right]dt \\ &+\left[\sqrt{\boldsymbol{G}\left\{\vartheta\left(t\right)\right\}}d\boldsymbol{B}\left(t\right)\right]_i.\end{aligned} \tag{31}$$

The discrete form of the stochastic differential equation provides a proposal as follows:

$$
\begin{aligned}
\tilde{\vartheta}_i = & \vartheta_i^o + \tfrac{\varepsilon^2}{2}\left\{\boldsymbol{G}^{-1}\left(\vartheta^o\right)\nabla_\vartheta L\left(\vartheta^o\right)\right\}_i - \varepsilon^2 \sum_{j=1}^{K_\vartheta}\left\{\boldsymbol{G}^{-1}\left(\vartheta^o\right)\frac{\partial \boldsymbol{G}\left(\vartheta^o\right)}{\partial \vartheta_j}\boldsymbol{G}^{-1}\left(\vartheta^o\right)\right\}_{ij} \\
& + \tfrac{\varepsilon^2}{2}\sum_{j=1}^{K_\vartheta}\left\{\boldsymbol{G}^{-1}\left(\vartheta^o\right)\right\}_{ij}\operatorname{tr}\left\{\boldsymbol{G}^{-1}\left(\vartheta^o\right)\frac{\partial \boldsymbol{G}\left(\vartheta^o\right)}{\partial \vartheta_j}\right\} + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}\left(\vartheta^o\right)}\boldsymbol{\xi}^o\right\}_i \\
= & \boldsymbol{\mu}\left(\vartheta^o, \varepsilon\right)_i + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}\left(\vartheta^o\right)}\boldsymbol{\xi}^o\right\}_i,
\end{aligned} \tag{32}
$$

where ϑ^o is the current draw. The proposal density is

$$
q\left(\tilde{\vartheta} \mid \vartheta^o\right) = \mathcal{N}_{K_\vartheta}\left(\tilde{\vartheta}, \varepsilon^2 \boldsymbol{G}^{-1}\left(\vartheta^o\right)\right) \tag{33}
$$

and convergence to the invariant distribution is ensured by using the standard form Metropolis-Hastings probability

$$
\min\left\{1, \frac{p\left(\tilde{\vartheta} \mid \cdot, \boldsymbol{X}\right) q\left(\vartheta^o \mid \tilde{\vartheta}\right)}{p\left(\vartheta^o \mid \cdot, \boldsymbol{X}\right) q\left(\tilde{\vartheta} \mid \vartheta^o\right)}\right\}. \tag{34}
$$

7 Appendix D: Posterior Analysis for Reduced Form Model

We use a Gibbs sampling algorithm that produces samples that converge in distribution to the posterior distribution of the model. This is achieved by simulating from the conditional distributions of parameters vectors given the rest of the parameters. We treat the inefficiency terms as a latent variables, which are considered to be parameters. In this section, for the sake of notational simplicity, we drop time index. The posterior augmented with inefficiencies for reduced form model is given by:

$$\begin{aligned}
&\left(2\pi\sigma_v^2\right)^{-n/2}\exp\left\{-\frac{1}{\sigma_v^2}\sum\nolimits_{i=1}^{n}\left(y_i - x_i'\beta - u_i\right)^2\right\}\cdot \\
&\left(2\pi\sigma_u^2\right)^{-n/2}\left\{\prod\nolimits_i^n \Phi\left(\frac{z_i'\gamma}{\sigma_u}\right)^{-1}\right\}\cdot\exp\left\{-\frac{1}{\sigma_u^2}\sum\nolimits_{i=1}^{n}\left(u_i - z_i'\gamma\right)^2\right\}\cdot \\
&\left\{\sigma_v^{-1}\sigma_u^{-1}\right\},
\end{aligned} \tag{35}$$

where $y_i = \ln C_i^*$ and $x_i'\beta = \ln C_i$. We will employ the following reparameterizations:

$$\begin{aligned}
h &= \frac{1}{\sigma_u} \\
\delta &= \gamma h.
\end{aligned} \tag{36}$$

The Gibbs sampler consists of the following steps. First, we set the initial conditions for β, σ_v^2, h, δ, and u. Initial values for β are obtained from OLS estimates; initial values for σ_v^2 and σ_u^2 are set to s^2 where s^2 is the OLS residual variance; initial values of other parameters are $\delta = 0$; and $u = 0.5$. For $gd = 1, ..., GD$ where GD is the number of Gibbs draws, we do:

1. Draw $\beta|\sigma_v^2, \sigma_u^2, \delta, u, y, X$
2. Draw$\sigma_v^2|\beta, \sigma_u^2, \delta, u, y, X$
3. Draw $h|\beta, \sigma_v^2, \delta, u, y, X$

4. Draw $u|\beta, \sigma_v^2, \sigma_u^2, \delta, y, X$
5. Draw $\delta|\beta, \sigma_v^2, \sigma_u^2, u, y, X$

These conditional posteriors are detailed below.

7.1 Drawing β

For β we have:

$$\beta|\cdot \sim \mathcal{N}_k(b, W), \tag{37}$$

where $b = (X'X)^{-1} X' (y - u)$ and $W = \sigma_v^2 (X'X)^{-1}$.

7.2 Drawing σ_v^2

For σ_v^2 we have:

$$\frac{Q_v}{\sigma_v^2}|\cdot \sim \chi_n^2, \tag{38}$$

where $Q_v = (y - u - X\beta)' (y - u - X\beta)$. We generate this as $\sigma_v^2 = \frac{Q_v}{d}$, where $d \sim \chi_n^2$.

7.3 Drawing σ_u^2

Note that we use the change of variable $h = \frac{1}{\sigma_u}$. The conditional posterior of h is:

$$p(h|\cdot) \propto h^{n-1} \exp\left\{-\frac{1}{2}(hu - Z\delta)'(hu - Z\delta)\right\}. \tag{39}$$

It can be showed that this function is log-concave. Suppose we wish to draw from a source which is a gamma density:

$$g\left(h,\alpha\right)=\frac{\alpha^{n}}{\Gamma\left(n\right)}h^{n-1}\exp\left(-\alpha h\right) \tag{40}$$

and seek the optimal value of α. Consider the ratio:

$$R\left(h,\alpha\right)=\frac{p\left(h|\cdot\right)}{g\left(h,\alpha\right)}\propto\alpha^{-n}\exp\left\{\alpha h-\frac{1}{2}\left(hu-Z\delta\right)'\left(hu-Z\delta\right)\right\}. \tag{41}$$

Now, let (α^{*},h^{*}) be the solution of saddle-point problem:

$$\min_{\alpha}\max_{h}\ln R\left(h,\alpha\right). \tag{42}$$

Then, we use acceptance sampling procedure as follows:

1. Draw $h\sim g\left(n,\alpha^{*}\right)$.

2. Accept the draw if $\frac{R(h,\alpha^{*})}{R(h^{*},\alpha^{*})}\geq U$ where U is a draw from standard uniform distribution; else go to step (1).

7.4 Drawing u

We draw each u_i independently from:

$$u_i|\cdot\sim\mathcal{N}_{+}(\hat{u}_i,\omega^2), \tag{43}$$

where $\hat{u}_i = \frac{\sigma_u^2 r_i + \sigma_v^2 q_i}{\sigma_v^2 + \sigma_u^2}$ and $\omega^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}$. To obtain $\hat{u}_i$ we combined two equations:

$$k_i = u_i + \xi_{i1}; \xi_{i1} \sim \mathcal{N}\left(0, \sigma_v^2\right); k_i := y_i - x_i'\beta, \tag{44}$$

$$q_i = \gamma_1 u_{-i} + z_i'\gamma_2 = u_i + \xi_{i2}; \xi_{i2} \sim \mathcal{N}\left(0, \sigma_u^2\right).$$

7.5 Drawing δ

The conditional posterior of δ is given by:

$$p(\delta|\cdot) \propto \prod\nolimits_i^n \Phi\left(z_i'\delta\right)^{-1} \cdot \exp\left\{-\frac{1}{2}(hu - Z\delta)'(hu - Z\delta)\right\}. \tag{45}$$

As this function is log-concave we can employ acceptance sampling as before:

1. Draw from a multivariate normal: $\delta \sim \mathcal{N}(d, W)$ with density $f(\delta; d, W)$.

2. Accept the draw if $\frac{p(\delta|\cdot)/f(\delta;d,W)}{p(d|\cdot)/f(d;d,W)} \geq U$ where U is a draw from standard uniform distribution; else go to step (1).

We determine d and W from the first and second derivatives of the following expression:

$$\ln p(\delta|\cdot) = -\sum\nolimits_{i=1}^n \Phi\left(z_i'\delta\right) - \frac{1}{2}(hu - Z\delta)'(hu - Z\delta). \tag{46}$$

In particular, we have:

$$\nabla \ln p(\delta|\cdot) = -\sum\nolimits_{i=1}^n \Lambda\left(z_i'\delta\right) z_i + Z'(hu - Z\delta) = 0 \tag{47}$$

$$\nabla^2 \ln p(\delta|\cdot) = -\left(\sum\nolimits_{i=1}^n \Lambda'\left(z_i'\delta\right) z_i z_i' + Z'Z\right), \tag{48}$$

where $\Lambda(x) = \frac{\phi(x)}{\Phi(x)}$ and Λ' is the derivative of Λ. We set d to be the solution of Equation (47) and $W = \left(\sum_{i=1}^{n} \Lambda'(z_i' d) z_i z_i' + Z'Z\right)^{-1}$.

8 Appendix E: Simultaneous Model Estimation Details

Let $v_{it}(\vartheta)$ denote the errors from System (11) and $\vartheta \in R^d$ the parameter vector. Then, the likelihood is given by:

$$L(\vartheta, \Sigma; \mathcal{Y}) = (2\pi)^{-kN/2} |\Sigma|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it}(\vartheta)'^{-1} v_{it}(\vartheta)\right\} \prod_{i=1}^{n} \prod_{t=1}^{T} J_{it}(\vartheta), \tag{49}$$

where $N = nT$ is the number of observations, $\mathcal{Y}$ is the data, $k = 3$, and $J_{it}(\vartheta) = ||\frac{\partial v_{it}(\vartheta)}{\partial \mathcal{Y}_{it}'}||$ is the Jacobian term. The error terms of the system $v_{it} = [v_{1,it}, v_{2,it}, v_{3,it}]' \sim N(0, \Sigma)$ and Σ is integrated out analytically using standard operations (e.g. Zellner, 1971, p. 385) to obtain:

$$L(\vartheta; \mathcal{Y}) \propto |\mathbf{A}(\vartheta)|^{-N/2} \prod_{i=1}^{n} \prod_{t=1}^{T} J_{it}(\vartheta), \tag{50}$$

where $\mathbf{A}(\vartheta) = \sum_{i=1}^{n} \sum_{t=1}^{T} v_{it}(\vartheta) v_{it}'(\vartheta)$. In System (11), instead of usual term $b_{it,0} + b_{it,1}t + b_{it,2}t^2$ used by Cornwell et al. (1990), we attach a term of the form $\exp\left(-\left(b_{it,0} + b_{it,1}t + b_{it,2}t^2\right)^2\right)$ and measure any additional non-structural inefficiency that may exist. This term is non-negative and less than one. Being smaller than one is not a priori necessary but useful in practice based on what we know about banking inefficiency. This retains the spirit of within estimator of Cornwell et al. (1990), possibly improves over them and permits the independent inclusion of country or bank fixed effects. Then, we apply a MCMC procedure as described in Appendix D using flat priors for all parameters, $p(\vartheta) \propto const.$, so that the posterior $p(\vartheta \mid Y) \propto L(\vartheta; Y)p(\vartheta)$ is proportional to the likelihood. We use $15,000$ iterations the first $5,000$ of which are discarded to mitigate possible start-up and convergence effects.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: