Is getting larger a problem? Evidence from European banks

M. Hasannasab,1,2 D. Margaritis,1,2 D. Mayes,2 C. Staikouras3
2University of Auckland Business School
3Athens University of Business & Economics

1. Introduction

Returns to scale (RTS) and their associated scale elasticity (SE) metrics are essential features of the production technology and play an important role in decision making, e.g. decisions regarding M&A targets or the premium an acquirer is willing to pay for a target. RTS is also an important input in competition policy since market power has traditionally been rooted in underlying scale economies in an industry. In banking, there are extra regulatory concerns about increased industry concentration insofar this encourages excessive risk taking as a result of implicit government guarantees for very large ‘too-big-to-fail’ (TBTF) banks, in addition to moral hazard problems arising from the presence of explicit guarantees, e.g. deposit insurance. Where the costs of achieving scale may be extensive, the losses of plurality arising from M&As2 or more generally from large banks getting bigger, are likely to be significant and potentially irreplaceable. This is particularly true if expansion is driven by high leverage which may not be sustainable, e.g. as a result of asset price run-ups, or from rapid expansion of loosely regulated off-balance-sheet (OBS) activities involving complex financial instruments which may prove excessively risky as evident from the subprime crisis. Deutche Bank’s daunting problems resulting from an over ambitious off-balance-sheet expansion, a very costly gamble, at a time other European banks were retrenching in the face of slackening markets and tightening regulation also come to mind. However, large banks may be getting bigger and historically have done so for reasons that may have more to do with potentially significant technological or cost advantages brought about by

1 Corresponding author: D. Margaritis, University of Auckland Business School, 12 Grafton Rd, Auckland, New Zealand. Email: d.margaritis@auckland.ac.nz
2 For example, Molyneaux, Schaeck and Zhou (2014) report that safety related implicit subsidies derived from bank M&A activity in Europe between 1997 and 2007 have a positive association with rescue probability, indicative of significant moral hazard problems. Davies and Tracey (2014) find no evidence of scale economies once they control for implicit subsidies associated with the reduction of funding costs for TBTF banks.
economies of scale. The intent of this paper is to examine if indeed such benefits exist and whether they are more common amongst the largest banks.

While TBTF banks have been under regulatory scrutiny for a long time, the recent global financial crisis (GFC) has widened the focus and importance attached to these institutions recognising that TBTF banks are likely to pose even greater risks to systemic stability in the event of a failure than previously thought. With concerns over large banks needing assistance looming, European lawmakers have been debating proposals giving regulators extra powers to cap bank size or split-up banks, in addition to tighter regulations on capital, leverage etc. introduced as a response to the financial crisis. Capping the size of banks may, of course, have a serious downside (see Wheelock and Wilson, 2016) if it prevents banks from exploiting economies of scale. Yet size is widely regarded as the main underlying cause of systemic risk and it is the benchmark generally adopted for the purposes of identifying systemically important financial institutions (SIFIs).

Using a sample of European banks, we examine if getting bigger is actually a cause for concern. We assess bank performance using a measure of bank efficiency constructed as the distance to the technological frontier, and investigate whether larger banks are exploiting technologically driven scale economies using a new measure of RTS based on closeness to the most productive scale size (MPSS). We pay particular attention to the relationship between bank size and bank performance and to the relationship between bank size and scale elasticity. Given the importance of additional systemic risk factors sprang from the GFC, such as liquidity problems stemming from excessive leverage and over ambitious OBS expansions, we also examine the relation between scale elasticity and liquidity ratio, leverage ratio, net interest margin, and investment securities to total assets ratio.

With the European Central Bank (ECB) assuming supervisory and regulatory responsibilities for SIFIs, mainly banks, the banking landscape across the European Union (EU) continues to evolve as regulators and policy makers debate the types of banking structures that best support economic growth. Naturally, informed debate and informed decisions require robust quantifiable evidence. This brings into the forefront the issue of RTS and SE measurement recognising the
difficulties researchers face to identify and quantify scale economies evident from the conflicting empirical findings reported in the literature. While studies generally find evidence of scale economies for large US banks in the 1990s and 2000s (e.g. see Berger and Mester, 1997; Hughes, Mester and Moon, 2001; Feng and Serletis, 2010; Hughes and Mester, 2013; Wheelock and Wilson, 2012, 2016) the evidence from European banks is quite mixed. For example, Altunbas, Gardener, Molyneaux and Moore (2001), Vander Vennet (2002), Merciera, Schaeck and Wolfe (2007) find scale economies for smaller European banks and generally diseconomies or constant returns to scale (CRTS) for larger banks. Dikstra (2013) reports scale economies across all bank sizes and so do Beccalli, Anolli and Borello (2015) although they find that economies are significantly greater for the largest banks in their sample.

We focus on approaches to RTS characterisation and SE measurement that are directly applicable to non-parametric multiple-output technologies (e.g. see Panzar and Willig, 1977; Färe, Grosskopf and Lovell, 1985; Banker and Thrall, 1992; Førsund, 1996; Sueyoshi, 1999; Fukuyama, 2000; Podinovski, Førsund and Krivonozhko, 2009; Balk, Färe and Karagiannis, 2015). Scale elasticity is thus defined in terms of distance functions (a generalisation of the more familiar production function) and is computed from Farrell (1957) technical efficiency measures and the multiplier (shadow value) of the convexity constraint of the dual data envelopment analysis (DEA) model. While DEA has been extensively used in the measurement of efficiency and productivity in banking, RTS measures have generally been obtained from parametric cost functions offering the convenience of smooth differentiable functional forms. One issue, however, with cost based measures is that prices for bank inputs are not readily available with questions being raised about the reliability of proxies used in empirical studies. A second issue is that RTS measures are meaningful for frontier points or for the projection of an inefficient point to the frontier but not for the inefficient point itself, since there is confounding between inefficiency and economies of

---

3 Inanoglu, Jacobs, Liu and Sickles (2012) and Davies and Tracey (2014) are notable exemptions.
4 Scale elasticity and economies of scale are related but different concepts with scale elasticity focusing on the relation between input and output quantities whereas economies of scale show the effect of an increase in output levels on unit costs. Whether one should focus on cost rather than technologically-driven economies is an open question recognising that cost advantages may include both the benefits of economies of scale as well as those of the implicit TBTF funding cost subsidies.
scale. This is a subtle point often not recognised in the literature where parametric stochastic frontier models lump together measures of scale economies for both frontier and interior points.

Previous methods of calculating scale elasticity for frontier points based on non-parametric piecewise linear technologies suffer from inherent ambiguities in choosing a single candidate as the final SE value. For example, SE may not be uniquely determined between an upper or lower bound or a particular value from the SE interval unambiguously chosen since the efficient point may lie on several efficient facets. Further complications arise, as the chosen value may be unrealistic, very sensitive to even small changes in input or output values as well as sensitive to the orientation of the DEA model.

In this paper, we use a new metric recently developed by Hasannasab, Roshdi, Margaritis and Rouse (2016) where SE is measured as the distance from the most productive scale size (MPSS) frontier. Calculating SE against MPSS gives a better economic meaning to the concept of SE since MPSS corresponds to the point on the efficient frontier with the maximum average productivity for the given input-output vectors (see Banker, 1984). Our approach also overcomes the limitations of previous methods offering a single meaningful numerical value for SE that provides the RTS characterisation directly, it is not sensitive to changes in inputs and outputs or the orientation of the DEA model, and is consistent across the solvers we use to solve the proposed model. Similar to previous studies our focus is on variable returns-to-scale (VRS) technologies, specifically on the frontier of the production possibilities set ($T_{VRS}$) that can be characterised by its supporting hyperplanes.

The main questions we set out to assess empirically are as follows:

- Are larger banks more efficient?
- Do larger banks enjoy higher returns to scale?
- Are returns to scale greater for more liquid and less leveraged banks?

The last question is part of series of questions we ask in an endeavour to establish whether some of the SE results stem from banks that have pushed themselves into fragile positions: because they are illiquid, highly leveraged, unprofitable or heavily focused on proprietary trading, which
has more recently come under control, with the Volcker Rule in the Dodd-Frank Act, requiring separation of such activities.

The paper is organised as follows. In Section 2 we introduce the relevant technical material on scale elasticity. Section 3 describes our approach to measure SE while Section 4 reports the empirical results. Section 5 concludes the paper.

2. Scale Elasticity and DEA

Assume there is a production process transforming a vector of \( m \) inputs, \( x = (x_1, ..., x_m)^T \in \mathbb{R}_{\geq 0}^m \), into a vector of \( s \) outputs, \( y = (y_1, ..., y_s)^T \in \mathbb{R}_{\geq 0}^s \). A convenient way to describe the production technology of this multi-input, multi-output process is to use the technology set that comprises all possible input-output vectors as

\[
\mathcal{T} = \{(x, y) \in \mathbb{R}_{\geq 0}^{m+s} | \ y \ can \ be \ produced \ by \ x\}.
\] (1)

By imposing some regularity conditions on \( \mathcal{T} \) (see Panzar and Willig, 1977), there exists a differentiable implicit function \( F(x, y) \), called a production transformation function, that alternatively characterises the technology, i.e.,

\[
F(x, y) \geq 0 \iff (x, y) \in \mathcal{T}.
\] (2)

For a single-output production process, this function reduces to the familiar notion of the production function, \( y = f(x) \), where \( y \) is the maximal producible value for every \( x \).

To measure SE we need to find the maximum proportional change in outputs that keep the transformation function equal to zero at every frontier point \( (x,y) \) of \( \mathcal{T} \) (i.e. \( F(x,y) = 0 \)) when inputs change proportionally (see Panzar and Willig, 1977). Hence SE can be directly calculated from the derivatives of the production transformation function as:

---

5 The superscript \( T \) stands for a vector transpose and bold letter represents a vector.

6 Exploiting the relationship between the transformation function and Farrell’s input and output distance functions, Färe et al. (1985) derived analogous formula for measuring SE in terms of input and output distance functions.
\[
\mathcal{E}(\mathbf{x}, \mathbf{y}) = - \frac{\sum_{i=1}^{m} \frac{\partial F(x,y)}{\partial x_i} x_i}{\sum_{r=1}^{s} \frac{\partial F(x,y)}{\partial y_r} y_r}
\]  

It can be shown that \( \mathcal{E}(\mathbf{x}, \mathbf{y}) \) is the marginal change in outputs scaling factor allowed by a marginal change in the inputs scaling factor over the average ratio of these factors, which reduces to the ratio of marginal productivity to average productivity for the simple single-input, single-output case. Clearly (3) is not valid at points inside the frontier even if this has not discouraged authors from reporting RTS results for inefficient firms. To operationalise our procedure we need to show the equivalence between marginal productivities in (3) and the dual variables (multiplier form) of a DEA program used to construct empirically the technology frontier.

Following Banker et al. (1984) the production possibilities set under the variable returns to scale DEA technology is formulated by imposing convexity, monotonicity, and minimum extrapolation postulates as

\[
\mathcal{T}_{VRS} = \{(x, y) \in \mathbb{R}_{\geq 0}^m \times \mathbb{R}_{\geq 0}^s : \ x \geq \sum_{j=1}^{n} \lambda^j x^j, \ y \leq \sum_{j=1}^{n} \lambda^j y^j, \ \sum_{j=1}^{n} \lambda_j = 1\},
\]  

in which \((x^j, y^j)^T, \ j = 1, ..., n,\) denote the input-output vector of the \(jth\) observation, generically referred as decision-making units (DMUs) in the DEA terminology.

It is well known that \( T_{VRS} \) is a polyhedral set and (4) exhibits its primal representation. The dual representation of this polyhedron is also possible in which \( T_{VRS} \) is described in terms of its supporting hyperplanes (e.g., see Briec and Leleu, 2003). There is also a relationship between the normal vectors of supporting hyperplanes of \( T_{VRS} \) at an efficient point and the optimal solutions of the input and output-oriented multiplier form of VRS models, presented in (6) below. These models were introduced by Banker (1980) to evaluate the relative efficiency of a given production point \((\mathbf{x}, \mathbf{y})\) operating under VRS:

---

7 See Førsund and Hjalmarsøn (2004) for more details.
\[
\begin{align*}
\text{max } & \mathbf{u}^T \mathbf{y} + u_o & (IM) & \text{min } & \mathbf{v}^T \mathbf{x} - u_o & (OM) \\
\text{s.t. } & \mathbf{u}^T \mathbf{y}^j - \mathbf{v}^T \mathbf{x}^j + u_o \leq 0, & & \text{s.t. } & \mathbf{u}^T \mathbf{y}^j - \mathbf{v}^T \mathbf{x}^j + u_o \leq 0, \\
& \mathbf{v}^T \mathbf{x} = 1, & & & \mathbf{v}^T \mathbf{x} = 1, \\
& \mathbf{u} \geq 0, \mathbf{v} \geq 0, & & & \mathbf{u} \geq 0, \mathbf{v} \geq 0, \\
& u_o \text{ unrestricted}, & & & u_o \text{ unrestricted}
\end{align*}
\]

Note that \( \mathbf{u} \) and \( \mathbf{v} \) are the output and input multipliers (shadow prices), respectively, and \( u_o \) is the multiplier (shadow value) of the convexity constraint, \( \sum_{j=1}^{n} \lambda_j = 1 \), characterising RTS. Hasannasab et al. (2016) show that the normal vector of every supporting hyperplane of \( \mathcal{T}_{VRS} \) at a given efficient point can be derived from the optimal solutions of these models.

Let \( (\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o) \) be an optimal solution of Model (IM) or (OM) at a given efficient point \( (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \). Since \( (\bar{\mathbf{u}}, -\bar{\mathbf{v}}) \) specifies the normal vector of a supporting hyperplane of \( \mathcal{T}_{VRS} \) at \( (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \), it can be utilized to locally describe the technology transformation function as

\[
F(\mathbf{x}, \mathbf{y}) := \bar{\mathbf{u}}^T \mathbf{y} - \bar{\mathbf{v}}^T \mathbf{x} + \bar{u}_o,
\]

where \( F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = 0 \). Implementing the general rule for this specific linear transformation into (3), yields the following formula for SE in a DEA setting

\[
\mathcal{E}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = - \frac{\sum_{i=1}^{m} \frac{\partial F}{\partial x_i} \xi_i}{\sum_{r=1}^{s} \frac{\partial F}{\partial y_r} \nu_r} \bigg|_{(\bar{\mathbf{x}}, \bar{\mathbf{y}})} = \frac{\bar{\mathbf{v}}^T \bar{\mathbf{x}}}{\bar{\mathbf{u}}^T \bar{\mathbf{y}}}.
\]  

Since DEA offers a piecewise linear frontier, rather than an everywhere differentiable function, we may face multiple supporting hyperplanes at some frontier units. If the given point, \( (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \), is a relative interior of the intersection of a supporting hyperplane of \( \mathcal{T}_{VRS} \) with the technology set, \( \mathcal{T}_{VRS} \), then the normal vector of the supporting hyperplane defines a unique value for SE via (7). To accommodate the case of non-unique supporting hyperplanes Banker and Thrall (1992) proposed finding the upper and lower bounds for SE by applying models (IM) and (OM). For this, they fix the optimal objective of (IM) at unity by adding the constraints \( \mathbf{u}^T \mathbf{y} + u_o = 1 \) to (IM), and then solve for both \( \min u_o \) and \( \max u_o \) as the optimal solutions of (IM). Let \( u_o^{-\text{IN}} \) and \( u_o^{+\text{IN}} \) be the minimum and maximum values of \( u_o \), respectively. For measuring SE

---

\footnote{This constraint ensures that the DMU under evaluation is located on the efficient frontier.}
pertaining to every optimal solution of Model (IM), given that \( \mathbf{v}^T \mathbf{x} = 1 \), expression (7) can be simplified as

\[
\mathcal{E}_{IN} = \frac{1}{1 - \bar{u}_o}.
\]

(8)

Now by applying (9) for \( u_{o,IN}^- \) and \( u_{o,IN}^+ \), the input-oriented interval of SE is identified as \([\mathcal{E}_{IN}^-, \mathcal{E}_{IN}^+]\).

Implementing the same procedure for the output-oriented model (OM), it can be shown that for every optimal solution of (OM), (8) reduces to the following formula

\[
\mathcal{E}_{OUT} = 1 + \bar{u}_o.
\]

(9)

Similarly, one can compute the upper and lower values of \( u_o \) from all optimal solutions of Model (OM), say \( u_{o,OUT}^- \) and \( u_{o,OUT}^+ \), and derive an output-oriented SE interval as \([\mathcal{E}_{OUT}^-, \mathcal{E}_{OUT}^+]\), by applying (9).

3. Most productive scale size and SE

Information about estimated MPSS allows decision makers to resolve questions about returns to scale for multiple inputs and output technologies without the need for additional information on prices or costs (see Banker et al., 2004). The MPSS for a given input and output mix is the scale size in which the outputs produced per unit of the inputs is maximised (see Banker, 1984). In other words, MPSS is a unit that maximises average productivity for its given input-output vector. If SE is greater than unity, this means the proportional changes in outputs are greater than the proportional changes in inputs so in order to increase average productivity one would increase the scale size. On the other hand, if SE for a given DMU is less than unity then the proportional changes in its outputs is less than proportional changes in its inputs hence average productivity can increase with a smaller scale size. Thus, in both cases, in order to reach a desired scale, the units are getting closer to the MPSS part of the frontier.

Cooper, Seiford and Tone (2007) demonstrate that if \((\bar{x}, \bar{y})\) lies on the CRS part of the technology, then it is an MPSS and thus there is a supporting hyperplane passing through it in the
form of \( \mathbf{u}^T \mathbf{y} - \mathbf{v}^T \mathbf{x} = 0 \) in which the value of \( u_o \) is zero. Equation (7) for this specific hyperplane implies that the SE value for any unit on this hyperplane is equal to one. Thus measuring SE with reference to the MPSS frontier reduces to finding a supporting hyperplane that offers an SE value close to unity, i.e., \( \mathbf{u}^T \mathbf{y} \) is as close as possible to \( \mathbf{v}^T \mathbf{x} \), or \( |\mathbf{u}^T \mathbf{y} - \mathbf{v}^T \mathbf{x}| \) is as close as possible to zero. Equivalently, \( |u_o| \) is as close as possible to zero.

Therefore, moving from each efficient point towards the supporting hyperplane that has a smaller absolute value of \( u_o \) will lead to an improvement in average productivity, i.e., the hyperplane that has a smaller absolute value of \( u_o \), is closer to the MPSS frontier of the technology. For any given efficient unit, the \( |u_o| \) value of the supporting hyperplane can thus be viewed as a divergence indicator from the MPSS-frontier. Hence we define our SE concept using the notion of “closeness” to the MPSS frontier as follows:

**Definition** Suppose \( \overline{H} = \{(x, y): \mathbf{u}^T \mathbf{y} - \mathbf{v}^T \mathbf{x} + u_o = 0 \} \) and \( \overline{H} = \{(x, y): \hat{\mathbf{u}}^T \mathbf{y} - \hat{\mathbf{v}}^T \mathbf{x} + \hat{u}_o = 0 \} \) are both supporting \( T_{VRS} \) at \((\overline{x}, \overline{y})\). We say \( \overline{H} \) is closer to the MPSS frontier than \( \overline{H} \) if \( |\overline{u}_o| \leq |u_o| \).

Let \((\overline{x}, \overline{y})\) be a given strong efficient point. SE can be measured at this point by identifying a supporting hyperplane that has a minimum distance from the MPSS frontier. To operationalize this idea, we introduce the following LP formulation:

\[
\begin{align*}
\text{max} & \quad \rho \\
\text{s.t.} & \quad \mathbf{u}^T \mathbf{y}^j - \mathbf{v}^T \mathbf{x}^j + u_o \leq 0, & j = 1, \ldots, n, \\
& \quad \mathbf{u}^T \overline{y} - \mathbf{v}^T \overline{x} + u_o = 0, \\
& \quad \rho \leq \mathbf{v}^T \overline{x} \leq 1, \\
& \quad \rho \leq \mathbf{u}^T \overline{y} \leq 1, \\
& \quad \mathbf{v}, \mathbf{u}, \rho \geq 0, \ u_o \text{ unrestricted}. 
\end{align*}
\]

(10)

The first set of \( n - 1 \) inequality constraints followed by the equality constraint for the efficient point \((\overline{x}, \overline{y})\) ensures that model (10) obtains a supporting hyperplane of \( T_{VRS} \) at \((\overline{x}, \overline{y})\). By choosing the optimal value of \( \rho \), the model maximises the weighted inputs and outputs values simultaneously thereby locating the supporting hyperplanes for which \( \mathbf{u}^T \overline{y} \) is as close as possible to \( \mathbf{v}^T \overline{x} \); and thus gives the closest hyperplane to the MPSS frontier.

Let \((\mathbf{u}^*, \mathbf{v}^*, u_o^*, \rho^*)\) be an optimal solution of Model (10). By using (7), the value of SE obtained
by Model (10) is

$$
\mathcal{E}_\rho = \frac{\mathbf{v}^\top \mathbf{x}}{\mathbf{u}^\top \mathbf{y}}
$$

(11)

Let \((\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{u}}_o)\) display the elements of a supporting hyperplane of \(T_{VRS}\) at the given strong efficient point \((\bar{x}, \bar{y})\). If \(u_o^* < 0\) \((u_o^* > 0)\), then \(\bar{u}_o < 0\) \((\bar{u}_o > 0)\) for any corresponding supporting hyperplane of \(T_{VRS}\) at \((\bar{x}, \bar{y})\). Alternatively, if \(u_o^* = 0\), then there is a supporting hyperplane of \(T_{VRS}\) at this unit so that \(\bar{u}_o = 0\). We formalise the relationship between \(u_o^*\) from Model (11) and \(\bar{u}_o\) from Model (IM) (or (OM)) as follows:

Assume \((\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{u}}_o)\) is an optimal solution of Model (IM). Then we have

(i) \(u_o^* < 0\) if and only if \(\bar{u}_o < 0\) for all optimal solutions of Model (IM);
(ii) \(u_o^* > 0\) if and only if \(\bar{u}_o > 0\) for all optimal solutions of Model (IM);
(iii) \(u_o^* = 0\) if and only if \(\bar{u}_o = 0\) for some optimal solution of Model (IM).

A similar result holds for Model (OM).

Model (11) classifies DMUs into strong efficient, efficient and inefficient points.

Remark \(\) At any point of \(T_{VRS}\) that is not a strong efficient point, Model (11) is feasible, bounded and has the value zero as its objective value, i.e., \(\rho^* = 0\). More precisely, assume \((\bar{x}, \bar{y})\) is a point of \(T_{VRS}\) and Model (11) has an optimal solution \((\mathbf{u}^*, \mathbf{v}^*, \mathbf{u}_o^*, \rho^*)\). Then one of the following will hold:

(i) \(\rho^* > 0\) if and only if \((\bar{x}, \bar{y})\) is a strong efficient point of \(T_{VRS}\).
(ii) \(\rho^* = 0\) and \((\mathbf{u}^*, \mathbf{v}^*) \neq 0\) if and only if \((\bar{x}, \bar{y})\) is an efficient point of \(T_{VRS}\) but it is not a strong efficient point.
(iii) \(\rho^* = 0\) and \((\mathbf{u}^*, \mathbf{v}^*, \mathbf{u}_o^*) = 0\) if and only if \((\bar{x}, \bar{y})\) is an interior point of \(T_{VRS}\).  

\(\)

4. Results

We use a sample of large European banks (assets > €30 billion). The data are obtained from Datastream and Bloomberg. We use as bank inputs: fixed assets, personnel expenses, total customer deposits and equity (as a proxy for bank risk). While the outputs are loans, other earning assets and off balance sheet assets. This choice of variables is highly contested in the

\(\)

Note that the result, viz. \(\rho^* = 0\), in (ii) and (iii) above is not restrictive since all DMUs can be projected to a strong efficient point along the frontier before SE can be calculated.
literature (Hughes and Mester, 2010). The primary debate is over what banks actually produce. While the basic process may be intermediation where they take funds, primarily deposits and short term, from one group of people and lend them to others at rather longer term, they clearly provide a much wider range of services. At the very least they offer, transactional facilities, risk mitigation instruments, custody, funds management, trade credit facilities, foreign exchange etc. The larger banks may also offer services related to new issues, insurance and many other areas of advice. While in the US, several of the activities may be separable from straightforward retail banking because of their holding-company structure, this tends not to be the case in Europe where a bank may be the apex entity in what is effectively a group.

There is therefore likely to be considerable heterogeneity among banks, which we reflect in our choice of ‘outputs’, considering securities and off-balance sheet assets. Still this does not give a measure of the level of activity in the banks in a particular period, just the consequences for the level of assets. Even if one could simply concentrate on loans, they are not homogeneous. They vary in their riskiness for a start. The rate of growth of the loan portfolio and its riskiness tend to be correlated.

We have made a pragmatic choice and seek to cover a range of coherent activities. This heterogeneity of activity is actually of benefit for our method. We employ a set of control variables to try to identify the conditions of banks that might affect their SE, such as leverage, liquidity, net interest margin and share of trading.

With these data we first construct the technology set, $T_{VRS}$, each year from 2004 to 2011 for the banks in the sample and then compute SE using the method outlined in Section 3. Efficiency results are shown across five quantiles of bank size in Figure 1 and across countries in Figure 2. Efficiency is highest in the Nordic countries and lowest in Austria and Cyprus. We find that the largest banks (Quantile 5) in our sample are the most efficient. However, it is interesting to note that until recently, small banks were more efficient than those of medium size, which came out worst – indicative of both a non-linear relation and in some respects a changing one.
In Figure 3 we show SE estimates for banks across the different countries (banking systems) in our sample. Finland seems to be an outlier\textsuperscript{10} whereas SE is lowest for banks in Belgium. We next proceed to study the relationship between SE and some of its underlying determinants. As determinants of SE we use bank size measured by total assets, liquidity measured as the ratio of liquid assets to total deposits, leverage measured as one minus the ratio of equity to total assets, profitability measured by the net interest margin, and business model measured by securities to total assets. Descriptive statistics are shown in Tables 1 and 2. Again we present our findings summarising our SE estimates across five quantiles sorted by the different bank characteristics.

Figure 4 shows the relationship between SE and bank size. We find an inverse relationship between SE and bank size. Smaller banks experience increasing returns to scale (IRTS) whereas larger banks face decreasing returns to scale (DRTS). This finding is contrary to the results by Beccalli et al. (2015) but in agreement with earlier studies on European banks. Moreover, if anything both the IRTS for small banks and the DRTS for larger banks are emphasised after the global financial crisis struck. We also find an inverse relationship between liquidity and SE. Figure 5 shows that SE is highest for the least liquid banks (Quantile 1) and lowest for the most liquid banks (Quantile 5). These findings are consistent with the results of Figure 4 insofar the largest banks happen to be the most liquid.

We find no obvious pattern between SE and bank leverage shown in Figure 6, with one exception. There is a major change in the most leveraged group at the height of the financial crisis, in part reflecting drastic measures to reduce leverage. There is also a negative association between net interest margin (NIM) and SE. Figure 7 shows that the more profitable banks appear to face DRTS. In relation to the business model, Figure 8 shows that banks with the lowest ratio of securities to total assets face IRTS whereas banks with a greater focus on investment banking activities face DRTS. This is an interesting finding given the concerns to protect banks from investment banking activities in the years since the global financial crisis struck. Overall our findings are indicative of

\textsuperscript{10} Pending a detailed investigation, this may reflect mergers, with Danske Bank buying Sampo Bank in 2006 and Nordea, the largest bank, moving from being headquartered in Finland to Sweden, thereby giving large fluctuations for what is nominally the same entity.
decreasing returns to scale (DRS) being quite prevalent in European banking systems with rather scant evidence that large European banks enjoy increasing returns to scale (IRS).

We proceed next to study scale elasticity in a multivariate context using panel regression methods. We report our results in Table 3. We find that more efficient, better capitalised and more liquid banks are associated with higher scale elasticity. It is of interest to note that some of these findings may be opposite in comparison to the simple bivariate results shown above highlighting the importance of studying multivariate relationships. We allow for a quadratic relationship between bank size and SE. For the majority of banks, the relationship is negative (there are only three banks with a positive relationship).

5. Conclusion

Scale elasticity is an important feature of production technology and plays an important role in decision-making, e.g. mergers and acquisitions, and in shaping competitive market structure and market conduct. Returns to scale play a particularly important role in banking since they not only they underlie M&A decisions but are also an input for decisions on bank regulation. Hence, it is important that SE is quantified correctly. While there have been several attempts to develop useful measures of SE, due to the linear piecewise nature of the DEA technology these have resulted in interval measures raising issues of indeterminacy and computational inconsistencies.

This paper has used a unique measure of SE that is based on a simple proposition of closeness to MPSS. The model is non-oriented and the result not only provides the SE measure but also allows returns to scale to be classified. The model was tested using a data set of 74 banks across 7 years. Overall, our findings are indicative of DRS in the European banking systems with very scant evidence that large European banks enjoy IRS. However, we find that on average the largest banks are the most efficient. Based on this evidence we surmise that although breaking-up Europe’s largest financial institutions may not be optimal concerns over capping their size may be exaggerated at least on the basis of the evidence we provide.
References


Figure 1: Efficiency across Bank Size Quantiles

Figure 2: Efficiency across countries
Figure 3: SE across Countries

Figure 4: SE across Bank Size Quantiles
Figure 5: SE across Liquidity Ratio Quintiles

Figure 6: SE across Bank Leverage Quantiles
Table 1: Descriptive statistics: Bank size quintiles (total assets in thousands of Euro).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.dev</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 - Smallest</td>
<td>32,869,067.8</td>
<td>12,136,985.77</td>
<td>8562,500</td>
<td>32,155,800</td>
<td>58,499,971</td>
</tr>
<tr>
<td>Q2 - Small</td>
<td>63,433,568.3</td>
<td>15,434,599.9</td>
<td>32,877,790</td>
<td>60,131,600</td>
<td>101,000,000</td>
</tr>
<tr>
<td>Q3 - Medium</td>
<td>135,550,484</td>
<td>50,016,942.22</td>
<td>40694,900</td>
<td>121,000,000</td>
<td>266,000,000</td>
</tr>
<tr>
<td>Q4 - Large</td>
<td>328,380,000</td>
<td>126,967,249.4</td>
<td>124,000,000</td>
<td>291,500,000</td>
<td>605,000,000</td>
</tr>
<tr>
<td>Q5 - Largest</td>
<td>1,088,247,423</td>
<td>469,538,027.2</td>
<td>445,000,000</td>
<td>957,000,000</td>
<td>2,200,000,000</td>
</tr>
</tbody>
</table>
Table 2: Descriptive statistics mean and St Dev [in brackets]

<table>
<thead>
<tr>
<th>Country</th>
<th>No. Banks</th>
<th>Total assets</th>
<th>Leverage</th>
<th>NIM</th>
<th>Sec/TA</th>
<th>Eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>2</td>
<td>138569785</td>
<td>0.9228</td>
<td>0.0436</td>
<td>0.30601</td>
<td>0.4575</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[61995089]</td>
<td>[0.0123]</td>
<td>[0.0104]</td>
<td>[0.0808]</td>
<td>[0.0286]</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
<td>282714285</td>
<td>0.9559</td>
<td>0.0302</td>
<td>0.45859</td>
<td>0.5832</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[25675632]</td>
<td>[0.0050]</td>
<td>[0.0050]</td>
<td>[0.0626]</td>
<td>[0.1819]</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1</td>
<td>31609157</td>
<td>0.9157</td>
<td>0.0328</td>
<td>0.28841</td>
<td>0.4693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[11114521]</td>
<td>[0.0387]</td>
<td>[0.0063]</td>
<td>[0.0604]</td>
<td>[0.1963]</td>
</tr>
<tr>
<td>Finland</td>
<td>2</td>
<td>124690857</td>
<td>0.9336</td>
<td>0.0204</td>
<td>0.58684</td>
<td>0.9271</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[118251294]</td>
<td>[0.0223]</td>
<td>[0.0049]</td>
<td>[0.0683]</td>
<td>[0.1370]</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
<td>465840364</td>
<td>0.9595</td>
<td>0.0289</td>
<td>0.51253</td>
<td>0.7872</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[547978178]</td>
<td>[0.0105]</td>
<td>[0.0219]</td>
<td>[0.2008]</td>
<td>[0.1690]</td>
</tr>
<tr>
<td>Germany</td>
<td>8</td>
<td>394408955</td>
<td>0.9621</td>
<td>0.0286</td>
<td>0.46816</td>
<td>0.8881</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[573963855]</td>
<td>[0.0129]</td>
<td>[0.0121]</td>
<td>[0.1796]</td>
<td>[0.1343]</td>
</tr>
<tr>
<td>Greece</td>
<td>4</td>
<td>67640907</td>
<td>0.9355</td>
<td>0.0430</td>
<td>0.23840</td>
<td>0.5690</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[24045412]</td>
<td>[0.0312]</td>
<td>[0.0090]</td>
<td>[0.0628]</td>
<td>[0.1534]</td>
</tr>
<tr>
<td>Ireland</td>
<td>4</td>
<td>97733591</td>
<td>0.9439</td>
<td>0.0248</td>
<td>0.32118</td>
<td>0.8295</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[62387458]</td>
<td>[0.0288]</td>
<td>[0.0086]</td>
<td>[0.1309]</td>
<td>[0.1402]</td>
</tr>
<tr>
<td>Italy</td>
<td>6</td>
<td>324664772</td>
<td>0.9294</td>
<td>0.0306</td>
<td>0.29602</td>
<td>0.7798</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[347814044]</td>
<td>[0.0194]</td>
<td>[0.0065]</td>
<td>[0.0628]</td>
<td>[0.1802]</td>
</tr>
<tr>
<td>Luxemburg</td>
<td>3</td>
<td>44949028</td>
<td>0.9640</td>
<td>0.0458</td>
<td>0.72123</td>
<td>0.8247</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[24514018]</td>
<td>[0.0205]</td>
<td>[0.0546]</td>
<td>[0.1242]</td>
<td>[0.2568]</td>
</tr>
<tr>
<td>Norway</td>
<td>2</td>
<td>115609660</td>
<td>0.9465</td>
<td>0.0190</td>
<td>0.17666</td>
<td>0.9185</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[71115487]</td>
<td>[0.0044]</td>
<td>[0.0016]</td>
<td>[0.0553]</td>
<td>[0.0767]</td>
</tr>
<tr>
<td>Portugal</td>
<td>3</td>
<td>90021396</td>
<td>0.9371</td>
<td>0.0251</td>
<td>0.26331</td>
<td>0.6409</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[19697285]</td>
<td>[0.0125]</td>
<td>[0.0035]</td>
<td>[0.0711]</td>
<td>[0.0685]</td>
</tr>
<tr>
<td>Spain</td>
<td>10</td>
<td>270955860</td>
<td>0.9415</td>
<td>0.0244</td>
<td>0.24766</td>
<td>0.7974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[351875338]</td>
<td>[0.0186]</td>
<td>[0.0081]</td>
<td>[0.0790]</td>
<td>[0.1328]</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
<td>222285714</td>
<td>0.9618</td>
<td>0.0147</td>
<td>0.36468</td>
<td>0.9649</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[29686642]</td>
<td>[0.0046]</td>
<td>[0.0011]</td>
<td>[0.0932]</td>
<td>[0.0680]</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2</td>
<td>888515453</td>
<td>0.9647</td>
<td>1.0138</td>
<td>0.70063</td>
<td>0.7453</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[407879157]</td>
<td>[0.0088]</td>
<td>[3.4335]</td>
<td>[0.1127]</td>
<td>[0.2533]</td>
</tr>
<tr>
<td>UK</td>
<td>13</td>
<td>527929966</td>
<td>0.9547</td>
<td>0.0261</td>
<td>0.41401</td>
<td>0.8544</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[548412567]</td>
<td>[0.0217]</td>
<td>[0.0107]</td>
<td>[0.2090]</td>
<td>[0.1825]</td>
</tr>
</tbody>
</table>
Table 3: Scale elasticity unbalanced panel regression with period fixed effects and cross-section (PCSE) adjusted standard errors (2005-2011)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOG(ASSETS)</td>
<td>-0.921006</td>
<td>0.187017</td>
<td>-4.924707</td>
<td>0.0000</td>
</tr>
<tr>
<td>LOG(ASSETS)^2</td>
<td>0.021440</td>
<td>0.004741</td>
<td>4.521890</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIER1</td>
<td>0.007108</td>
<td>0.003317</td>
<td>2.143094</td>
<td>0.0326</td>
</tr>
<tr>
<td>LIQRATIO</td>
<td>0.002960</td>
<td>0.001348</td>
<td>2.195905</td>
<td>0.0286</td>
</tr>
<tr>
<td>LLP/ASSETS</td>
<td>-0.830841</td>
<td>0.396912</td>
<td>-2.093265</td>
<td>0.0369</td>
</tr>
<tr>
<td>EFFICIENCY</td>
<td>0.219384</td>
<td>0.085322</td>
<td>2.571261</td>
<td>0.0104</td>
</tr>
<tr>
<td>C</td>
<td>10.43441</td>
<td>1.816023</td>
<td>5.745748</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared     | 0.289421    | Mean dependent variable | 0.955298     |
Adjusted R-squared | 0.271507   | S.D. dependent variable | 0.235780     |
S.E. of regression  | 0.201243   | Prob(F-statistic)       | 0.000000     |

LLP = loan loss provision