Market size and market structure in banking

Sara Alfaihani[1], Oleg Badunenko[2], Shabbar Jaffry[3]

1. Economics and Finance Department, College of Business Administration, University of Bahrain, P.O. Box 32038, Shakir, Bahrain. Email: salfaihani@uob.edu.bh
2. Department of Economics and Finance, College of Business, Arts and Social Sciences, Brunel University London, Uxbridge, UB8 3PH, UK. Email: oleg.badunenko@brunel.ac.uk
3. Economics and Finance Subject Group, Faculty of Business and Law, University of Portsmouth, Portsmouth, PO1 3DE, UK. Email: shabbar.jaffry@port.ac.uk

Abstract

Larger markets are usually characterised by a large number of firms. We show this does not hold in banking. Our analysis of banking in Gulf countries suggests the existence of a lower bound to concentration. The bounds are different across countries, however, in all of them, the markets remain concentrated regardless of the market size. We also show that the largest banks are becoming more dominant over time. Most importantly, the values of the limiting levels and the actual levels of concentration are astonishingly close suggesting that the banking markets of the Gulf countries are operating close to long-run equilibrium.

Keywords: Banking; Market Structure; Market Size; Concentration; GCC; Stochastic Frontier

JEL classification: G21, L1

1. Introduction

The relationship between the market size and market concentration is traditionally thought to be inversely related (see e.g. Dick, 2007). Larger markets are believed to be less concentrated than smaller ones, given that a higher number of firms is required to serve larger markets. Does this mean that the nature of competition among firms in larger countries should differ from that in smaller countries? Correspondingly, is it expected that the markets in larger countries are fragmented (i.e. low concentration)?

The discrimination between exogenous and endogenous sunk costs industries alters the traditional interpretation of the negative relationship between market size and market concentration. Sutton (1991)[1] is the first to propose the bound approach to model the size–structure relationship and distinguish the equilibrium structure between the two types of industries. Sutton's principal argument is that the inverse size–structure relationship is valid only for certain groups of industries, while it breaks down for industries in which shifting consumer demand relies on the fixed rather than variable costs. The empirical estimation of the relationship plots a lower bound to concentration. For exogenous sunk costs industry, the lower bound would be sharply decreasing and eventually approaches zero with limiting level of concentration below 1%. Such industries are of competitive nature where escalating profits due to market expansion induces entry to the market and hence lowers the concentration level. However, the estimated bound is remarkably flat for endogenous sunk costs industries. This

[1] The basis of Sutton's analysis was developed in two earlier papers by Shaked and Sutton (1987, 1983)

would suggest that number of entrants is entirely unaffected by the size of market and the industry is a natural oligopoly.

As far as theory goes, the entry to such industries is naturally limited to a handful of firms, when quality improvement and shifting the technological frontier outward is predominantly due to endogenously determined fixed costs. A continuous increase in the level of fixed costs is not always feasible for all firms. Therefore, it is believed that there exists a lower bound to the equilibrium level of concentration as the number of entrants is unrelated to the market size, and hence the market fails to fragment. Conversely, as the market size increases, the current incumbents will escalate their fixed investments. As a result, larger markets will not have a greater number of firms; instead the theory predicts that there will be an improvement in the quality of products.

Although Sutton's theory is developed in the early nineties and provides solid predictions of the size–structure relationship across a broad class of competition models, surprisingly, the empirical applications of Sutton's theory has been remarkably sparse. Ellickson (2007) is one of few exceptions. He was the first to test the theory on a large sample of markets within a single industry. His work examines the theory's predictions on the US supermarket industry, where a few firms were found to dominate the market and capture the majority of sales regardless of the market size. Also, the estimated lower bound was remarkably flat, indicating that the market concentration is unaffected by the market size.

The empirical testing of Sutton's theory requires defining a proxy of the set-up costs, which represent the minimal level of sunk cost required for anyone to enter the industry. However, this information is not available for most industries or is difficult to estimate. For instance, Dick (2007) applies Sutton framework to the US banking industry but the market size was not

scaled by the set-up ratio as it appears to be that there is no legal minimum. Inasmuch as the Sutton analysis is powerful to make prediction about the relationship between market size and market concentration, the empirical analysis requires data on entry requirements, which are not easy to come by. The limited empirical research testing of Sutton's theory owes to failure to obtain these data because it either does not exist at the micro level or cannot be disclosed due to confidentiality.

Once we use a panel data and thus deviate from the basic empirical Sutton cross-sectional model, we need to consider modifications. Over a longer time frame, the market structure need not remain the same. The question is not only whether the structure remains the same, but if it changes, does it become more or less concentrated. Also, the market may have segments, whose submarket structure vary. For example, some maybe an oligopoly, whole others natural monopoly. These two conjectures are subject to empirical test. Hence, we contribute to the literature first by proposing several empirical extensions to the baseline model of Sutton. In particular, we analyse the size–structure relationship in three different ways: (i) assume the market is the same for all countries, but time-varying, (ii) assume the markets are different, but are time-invariant, and (iii) assume both that the markets are different and are time-varying.

Empirically, the Sutton's (1991) model uses a cross-sectional data and applies either maximum likelihood or minimum distance estimator involving the error term following the extreme–value Weibull distribution. This study–to the best of our knowledge–is the first to attempt using panel data to test Sutton's theoretical size–structure model. The panel data models typically address the issue of unobserved heterogeneity that may be viewed as missing variables problem. It can also reflect time-invariant features of the observations which may or may not be related to already included variables into the model. There are different ways to account for the unobserved heterogeneity. For example, include as many time constant variables as

possible. Empirically though one could not be sure that the included variables sufficiently account for the unobserved heterogeneity. So usually, the so-called individual effects are included into the model. Both the minimum distance estimator, which can be viewed as a linear programming problem, and maximum likelihood estimator of the Weibull distribution-based model will be a challenging endeavour in this respect. In this paper, therefore, we make the second contribution by making use of the stochastic frontier approach in a panel data context to test the Sutton's theory over time and across countries. This technique is robust to outliers and allows low concentration disequilibria (Tabacco, 2013). We have tried several variants of the stochastic frontier for panel data and even though there are expected quantitative differences, they all deliver a consistent conclusion. As a robustness check, we also applied minimum distance estimator (assuming the data are a pooled cross-sectional data) and the conclusions remained qualitatively the same.

As we argue below, the proposed modelling of the size–structure relationship in a panel context has wide applicability as it generally allows addressing a number of important questions. *First*, does a lower bound to concentration exist in the industry? *Second*, is the industry of an exogenous or endogenous sunk-costs type? *Third*, is the limiting level of concentration changing across time? *Fourth*, is the nature of competition among firms in large countries should differ from that in smaller countries? *Fifth*, does the level of the equilibrium market structure vary across countries due to the set-up costs and size? *Finally*, are the values of the limiting and actual levels of concentration close or apart from one another?

In this study, we investigate the size–structure relationship of six Gulf Cooperation Council (GCC) banking markets that vary greatly in size thus providing a good case study. The findings of this paper indicate that there exists a lower bound to concentration in the banking industry of the GCC where the market structure remain concentrated across all market sizes. This is a

feature of endogenous sunk cost industries. As the market grows larger, the dominant banks will raise the provisions of quality through incurring additional fixed costs, which is not always feasible for all banks. This potentially makes entry for new banks harder. Therefore, it is the level of fixed costs instead of the number of banks that will escalate with market growth. Moreover, the limiting level of concentration have an upward trend over time, which in turn can imply that dominant banks have an increased market share over time. Though, the equilibrium levels of concentration vary across individual countries and are the highest for Qatar and Bahrain, whereas they are the lowest in the UAE and Saudi Arabia. This ranking is due to the level of set-up costs and market size. Interestingly, the banking markets of the GCC countries seems to be are operating under a long-run equilibrium as the limiting and actual levels of concentration are astonishingly close.

This paper is organised as follows. Sections 2 and 3 provide a review of the background of the GCC banking and the relevant literature on the market size and competition nexus. Section 4 describes the data, and the empirical methodology is explained in Section 5. Section 6 reports on the empirical results while Section 7 discusses the findings. Finally, Section 8 provides conclusions.

2. Background of the GCC

The GCC was established in 1981 among six oil-exporting countries: Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the UAE to achieve socio-economic development and cooperation that would eventually lead to greater economic growth and integration (Al-Muharrami et al. 2006). The GCC economies are highly dependent on the oil and gas sector and, for a long time, these countries have been following enormous diversification strategies. The recent fall of oil

prices[2] has reaffirmed the need for economic diversification. In the GCC, fluctuations in oil prices have a significant impact on financial stability.

Banks in the GCC play a central role in the national diversification plans as the lead finance providers for a broad range of sought-after investments in different sectors of the economy. Moreover, the percentage of banks' total assets to GDP is relatively high. The ownership structure of GCC banks is somewhat concentrated, either through government or family-group memberships (Ashraf et al., 2016). This explains the trivial share of foreign banks' branches in the GCC, which on average does not exceed 10% of the total banking assets.

The banking services in the GCC are offered by both conventional and Islamic banks. The latter operates in compliance with Sharia principles, which prohibit the payment or receipt of interest. In this manner, banks reward depositors with non-guaranteed returns and charge borrowers either a cost-plus or a profit share basis on loans. Islamic banks represent a significant share of the sector, representing one-quarter of the total banking assets in the Gulf region (Basu et al., 2018). With the exception of Bahrain, the majority of both Islamic and conventional banks in all the GCC countries are licensed to operate as retail banks. In the case of Bahrain, however, the banking structure is almost equally divided between retail and wholesale[3] banks. Given such special nature of GCC banking, our analysis focuses on retail banks and include both conventional and Islamic banks. In all GCC countries, the same regulations about the sunk cost is applied to both conventional and Islamic banks with exception to Oman. This is because Islamic banks were not allowed to operate in the country until late 2012.

[2] The West Texas Intermediate crude oil price reached a peak of over $150 per barrel in mid-2008 and hit a sharp downturn to $28 in early 2016.
[3] According to the Central Bank of Bahrain Rulebook (Volume 1-LR1.2.2), the wholesale banks category represents the offshore banking unit and investment banks.

As they gain in size, GCC banking markets are constantly becoming more important internationally. Figure 1 shows a timeline evolution of the size of GCC banking markets over the period 2000 to 2017. In the early 2000s, the banking markets of the GCC were in their infancy, with total banking assets below 300 billion US dollars in real terms. However, there was a rising trend over the period, and the markets have grown beyond US$1,200 billion in real terms. It can be clearly seen that the Saudi banking market was the largest until the UAE stepped forward in 2017. Hence, the Saudi and Emirati banking markets are the largest in the region. In contrast, Bahrain and Oman are the smallest markets, whilst Kuwait and Qatar are the mid-sized markets in the region.

Over the last three decades, the GCC banking markets went through significant and promising reforms. Primarily, these reforms introduced the implementation of financial liberalisation policies and financial restructuring policies. More recently, central banks in the GCC countries have introduced corporate governance-related standards and risk management rules in accordance with Basel III. The sole objective is to enhance the sector's competitiveness (Abuzayed et al., 2018). Banks do and will be playing a prominent role in GCC countries, so special attention needs to be paid to the market structure and concentration levels.

3. Literature Review

There is an extensive literature on market structure. One of the famous theories is the Structure–Conduct–Performance (SCP) paradigm of Mason (1939) and Bain (1956) that relates to cross-industry studies. The theory represents one-chain causation from the level of market concentration 'structure' to the extent of collusion 'conduct' and then to a firm's profitability

'performance'. Market concentration would enable the exploitation of firms' market power and ease collusive agreements to set higher prices and hence enjoy higher profitability. The SCP approach has received considerable criticism by both empirical researchers and contributors to the game theoretic literature. Nonetheless, this pioneering theory laid the foundations for a generation of subsequent work in the market structure literature.

Traditionally, the relationship between market size and the level of market concentration is deemed to be negative. The expansion of market size will increase the profits of incumbents, which will induce new firms to enter the market. This in turn will reduce the market share of every firm in the market and lead to a fall in market concentration. However, Sutton (1991) introduces a new theoretical rationale for this negative relationship by incorporating the exogenous and endogenous elements of fixed costs as possible explanations for a firm's conduct and market structure. Sutton (1991) argues that the inverse size–structure relationship is valid only for certain groups of industries, and the relationship breaks down for industries in which enhancing consumers' willingness to pay relies on sunk costs.

The basis of Sutton's analysis was developed in Shaked and Sutton (1987, 1982) and elaborated in Sutton (1991). The central notion of the theory lies in the interaction of endogenous and exogenous sunk costs with each other to determine the equilibrium industry structure. The level of sunk costs can be determined either exogenously (e.g. scale economies) or endogenously (e.g. advertising, and research and development (R&D) expenditures). In essence, sunk costs are captured through a two-stage game model of an industry equilibrium. At the first stage of the game, a firm incurs fixed outlays to acquire a single plant of the minimum efficient scale set-up costs and to establish a product line probably through advertising and R&D. The fixed outlays of the first stage are treated as sunk costs to analyse the nature of price competition in the second stage of the game.

In an industry with exogenous sunk costs, the level of concentration diminishes indefinitely as market size increases. An increase in market size will raise profits of firms and thus encourage further entry, which will lead to a fall in market concentration with market size expansion. This is in accordance with some well-known limit theorems of the economic theoretical literature, where it is maintained that scale economies become a trivial constraint on equilibrium structure in large economies. Within this type of industry, tougher price competition will also raise the equilibrium concentration level. The intuition behind this is that fierce competition among firms makes the entry less attractive to potential entrants and so raises the equilibrium levels of concentration. Hence, a rise in market size leads to a fragmented market and indefinitely low levels of concentration when sunk costs are determined exogenously. This property, nevertheless, breaks down for endogenous sunk-cost industries.

When sunk costs are determined endogenously, the level of concentration is relatively invariant to variations in market size. In such industries, it is endogenous sunk costs instead of the number of firms that will increase with the expansion of markets. Shaked and Sutton (1987) demonstrate that entry to certain industries is limited to a handful of firms when the burden of quality improvement and shifting the technological frontier outward falls predominantly on endogenous fixed costs. Such a continuous increase in the level of fixed costs is not always feasible for all firms. Therefore, it is believed that there exists a lower bound to the equilibrium level of concentration. In this case, the number of entrants is unrelated to market size, and hence the market fails to fragment. Conversely, as market size increases, the current incumbents will escalate their fixed investments. As a result, larger markets will not have a greater number of firms but rather will have better quality of products.

The econometric test of Sutton's theoretical framework is to estimate a lower bound to the market size–structure relationship and a limited value for the estimated bound 'C^{∞}'. Therefore,

the test depends on two key variables of market structure and a ratio of market size to set-up cost. Market structure is the observed level of concentration ratio measured by maximal market share of the single largest firm 'CR1'. Market size can be measured by the population, total assets, or sales volume of firms in the market. Set-up costs are defined as the minimum level of sunk costs that each entrant to the industry must make before commencing production.

The novelty of the theory is to distinguish the equilibrium structure between exogenous sunk-cost and endogenous sunk-cost industries. In industries where sunk costs are exogenously determined, there is a strong negative correlation between concentration and the ratio of market size to set-up costs. The minimal level of concentration approaches zero as the market size escalates. Contrariwise, these properties do not hold for industries where sunk costs are endogenously determined. This is attributed to two predictions. First, the minimal level of concentration is bounded away from zero, independent of market size. Secondly, the function describing the lower bound to concentration as a function of market size is not monotonic. Hence, endogenous fixed costs may represent an exogenous barrier to entry.

The properties of a specific industry will be the same for any country. Under general conditions, the equilibrium structure of an industry tends to be very similar across different countries. An industry that is dominated by few firms in one country is also likely to be dominated by few firms elsewhere. Many studies attribute this regularity across countries to the presence of the same pattern of technology and tastes in a specific industry, irrespective of which country is considered.

In one of the earliest applications of Sutton's theory, Ellickson (2007) shows that the supermarket industry is a natural oligopoly, where a handful of firms dominate the majority of sales regardless of the market size. The estimated lower bound is remarkably flat, and the level

of equilibrium concentration is asymptotic to positive levels. There is evidence that larger markets enjoy a greater variety of products in each store as firms escalate the fixed investments in firm-level distribution systems with market expansions.

Similarly, Dick (2007) finds that concentration remains similar across different sizes of markets in the banking industry. The quality of banking services is higher in larger markets and is mainly provided by dominant banks. The study suggests that when market size grows, banks accelerate fixed-cost quality investments to capture additional demand, and so entry barriers rise.

There remain several aspects of the size–structure relationship about which relatively little is known. In particular, the time and country heterogeneity have been neglected aspects of this relationship. The size–structure relationship can vary over time or in different countries. Moreover, the empirical estimations of this relationship are rather sparse in the literature. This study, therefore, tests the theory on the banking industry of the Gulf region. Unlike previous studies, we use panel data to provide extensions that capture the variations in the equilibrium level of concentration across time as well as countries. The next sections explain the data and methods.

4. Data

This study utilises a unique data set of GCC banks that includes all national retail banks that have operated during the period 2000–2017. We combine data from different sources to ensure the full coverage of each bank in the sample and to cross-validate the data. The banking data sources are annual reports of banks, Bloomberg, S&P Global Market Intelligence, and

Bankscope. Population data are obtained from the World Development Indicators provided by the World Bank.

Set-up costs are extracted manually from the banking regulations of each country. A comprehensive review of banking laws and bank licencing requirements in every GCC country, shows that all countries share similar explicit setup cost items. These are mainly in form of the minimum paid-up capital required to establish a bank, as well as a fixed amount for licensing fees. Therefore, we proxy the setup costs as the sum of the minimum paid-up capital required and the licensing fees.

It has become apparent, that the magnitude of these items can be broadly inferior, specifically when the minimum paid-up capital required in Saudi Arabia and UAE is compared to the other GCC countries. A possible explanation is that the banking regulations regarding a new bank establishment in other GCC countries have been amended in recent years. For example, the banking laws in Oman and Bahrain were amended in 2014 and 2017 respectively. In the case of Saudi Arabia and UAE, these laws remain unchanged since the 1960s.

Data on international purchasing power parity are exported from the University of Groningen Research database. This is used to report the total banking assets and set-up costs in real terms. Average US dollars exchange rates are from the International Financial Statistics provided by the International Monetary Fund. This is used to convert data from annual reports in local currencies into US dollars, for a unified currency within the sample.

Further, we apply three selection criteria to our sample to improve sample comprehensiveness and reliability and to avoid double counting (similar strategies are followed in the literature, e.g. Claessens and van Horen, 2014; Clerides et al., 2015). Firstly, the sample includes only

retail banks that carry out traditional intermediation services to ensure the consistency of bank types included in the sample. In the GCC context, retail banks comprise commercial (conventional) and Islamic (Sharia-compliant) banks as they both offer retail banking services to consumers but vary in the process if operations. We keep them in our sample as they both define the industry and sunk costs that are important in our analysis are applicable to either type of a bank. Also, subsidiaries of these banks are checked for and are included in the sample to prevent the false omission of some banks. In this vein, the sample excludes other types of banking institutions that have a financial structure less reliant on traditional deposit–lending activities; these are investment banks, wholesale banks (mainly in Bahrain), real estate and mortgage banks, and specialised governmental banks. The sample also excludes foreign banks given the minor market share and the scope of activities.

Secondly, we account for all mergers and acquisitions (M&As) that took place during the sample period. In the event of M&A, both banks appear in the sample before the M&A, whilst only the merged entity or the acquiring bank is included post-M&A. All separate bank accounts of the two merged banks or the acquired bank that were still reported for some years after the M&A are excluded from the sample. For example, National Bank of Dubai and Emirates International Bank merged to create Emirates NBD Bank in the year 2007. Thus, the two merged banks are included until 2006, and Emirates NBD bank appears from 2007 onwards. Appendix A gives the list of GCC banks in our sample and provides details of the M&A that occurred during the considered period.

The last criterion is to provide well-checked data and to include all national retail banks that have existed and operated during the last 18 years. Unlike other banking studies of the GCC, our sample includes *all* national-retail banks that are licensed by central banks in all of the six countries. This is confirmed through the list of banks provided by the central banks' reports.

This is a very crucial step to ensure the reliability of data, avoid any sample biases, and eventually to enrich the empirical results. Therefore, our sample represents an exceptionally reliable, homogenous, and comprehensive data set of GCC banks, which is rarely found in banking studies of the region.

Summary statistics of the sample are reported in **Error! Reference source not found.**. On average, Saudi Arabia represents the largest market in terms of population, assets, and revenues, while the UAE is the second largest market. Focusing on banking assets and revenues, Qatar and Kuwait are on average at the third and fourth rank, followed by Bahrain and Oman. In terms of population, however, the ranking differs. Oman and Kuwait are the largest markets by average population, whereas Qatar and Bahrain are the smallest. Moreover, the level of set-up costs is quite similar across GCC countries except for Saudi Arabia and the UAE, where they exhibit the minimum levels.

5. Methodology

5.1. Market Structure

To capture market structure, we use the two most popular measures: the *k*-banks concentration ratio (CR_k) and the Herfindahl–Hirschman Index (HHI).

The *k*-bank concentration ratio can be simply measured by summing the market shares of *k* banks in the market. Algebraically, it is measured as:

$$CR_k = \sum_{i=1}^{k} MS_i \qquad (1)$$

where MS is the market share of the *i*th bank and *k* is the number of the largest banks included in the calculations of the ratio. For all values of *k*, the index values range between zero and unity. Zero value of the index indicates an infinite number of equally sized banks, while unity is reached if the entire market is made up of the banks included in the calculations of the ratio. In other words, larger values of CR mean higher market concentration. A distinguishable aspect of the CR_k index is that it does not discriminate between the *k* leading banks.

Herfindahl (1950) and Hirschman (1964) developed the Herfindahl–Hirschman Index independently. The HHI Index has been the basis of merger guidelines in the US Department of Justice since 1982 and is widely used by bank regulatory agencies. The HHI is a static measure of concentration at a single point of time and is calculated by squaring the market share of each bank in a particular banking market and then summing the squares, as follows:

$$HHI = \sum_{i=1}^{n} (MS_i)^2 \qquad (2)$$

As mentioned above, MS refers to the market share of the *i*th firm, while *n* refers to the number of firms in the market. The HHI values range from 10,000 (in the case of a pure monopoly) to a number approaching zero (in the case of an atomistic market). According to the 2010 Horizontal Merger Guidelines, the markets are classified into three types: unconcentrated markets (HHI < 1500); moderately concentrated markets (1500 < HHI < 2500); and highly concentrated markets (HHI > 2500). In general, mergers in a moderately and highly concentrated market involving an increase in the HHI > 100 points are likely to raise substantial

competitive concerns and often warrant scrutiny (U.S. Department of Justice and the Federal Trade Commission, 2010).

5.2. The Lower Bound to Concentration

Sutton's theory is formally tested through the estimations of the lower bound of concentration when market size increases indefinitely. We follow Sutton (1991) and use the following specification:

$$\tilde{C}_{kit} = \beta_0 + \frac{\beta_1}{\ln(S_{it}/\sigma_i)} + \varepsilon_{it} \qquad (3)$$

where $\tilde{C}_{kit}$ is a logit transformation of the relevant concentration ratio for country *i* at time *t* and $\varepsilon_{it} > 0$. In our model, we consider two concentration ratios CR_1 and CR_3. So $\tilde{C}_1 = \ln(CR_1/(1 - CR_1))$ and $\tilde{C}_3 = \ln(CR_3/(1 - CR_3))$. This transformation is to remove the scale and ensure that the predicted values of concentration remain between 0 and 1. S_{it} corresponds to the market size of the banking industry measured by the total banking assets. The lower bound to concentration is $C^{\infty} = e^{\beta_0}/(1 + e^{\beta_0})$ and it depends only on the intercept in Eq. 3. For robustness, we also use population as another measure for banking market size. σ_i is a proxy of set-up cost that captures the minimal sunk cost required by each entrant to the industry prior to establishing production. In the GCC banking context, we refer to the regulations of the central banks and organisations of banking in each country and proxy the set-up costs to the sum of the minimum paid-up capital required and the licensing fees.

To econometrically estimate the lower bound of $\tilde{C}_{kit}$, a distributional assumption of the concentration measures and hence on ε_{it} needs to be made. Sutton and others follow the extreme-value Weibull distribution, since CR1 represents the greatest value in the distribution

of market shares observed in a sample and hence can be treated as an extreme value. This argument, however, applies somewhat loosely to larger CR_k (Ellickson, 2007). We adopt an alternative modelling strategy, which decomposes the error term ε_{it} into u_{it} and v_{it}. This allows making the Sutton model stochastic. Eq. 3 then becomes:

$$\tilde{C}_{kit} = \beta_0 + \frac{\beta_1}{\ln(S_{it}/\sigma_i)} + u_{it} + v_{it} \qquad (4)$$

In specification (4), $u_{it} > 0$, while v_{it} is the usual two-sided error term. The parameters in equation (4) can be estimated using the SF approach (see e.g. Ondrich and Ruggiero, 2001; Tsionas, 2017).[4] The one-sided error component u_{it} can be viewed as the deviation of the dependent variable from the potential minimum. To precede the potential concern that the error component may be correlated with S_{it}, we have run the SF model of Park et al., (1998), where effects maybe correlated with regressors. The conclusions that we will make using standard SF models are not different qualitatively from those using this more computationally involved SF model.[5]

The u_{it} error component can be assumed to follow different distributions such as half-normal (Aigner et al., 1977) or a more flexible two-parameter Gamma distribution (Greene, 1990), although some caveats are in order (Ritter and Simar, 1997). In practice, the distributional assumption does not make huge differences (Kumbhakar and Lovell, 2003), so we will assume that u_{it} follows a half-normal or truncated half-normal distribution.

While the baseline model of Sutton assumes that all countries are the same market, the market we are looking at is quite big and the time span is enormous to consider only a static model. In

[4] While the SF approach is typically used to evaluate performance of banks (see e.g. Dong et al., 2016), we propose to employ the SF approach for estimating the lower bound.
[5] The full set of results is available from authors upon request.

accordance with hypotheses put forward above, we suggest several extensions to account for both time and country heterogeneity on this relationship. In particular, we analyse market structure in GGC in three different ways: (i) assume the market is the same for all GCC countries, but time-varying (Extension 1), (ii) assume the markets are different and are time-invariant (Extension 2), and (iii) assume both that the markets are different and are time-varying (Extension 3). Next, we explain the empirical strategy to test these hypotheses.

In the first extension, to capture the disparities of the size–structure relationship and the equilibrium level of concentration over a period of time, we propose the following functional form that incorporates a nonlinear time trend t as:

$$\tilde{C}_{kit} = \beta_0 + \frac{\beta_1}{\ln(S_{it}/\sigma_i)} + \delta_1\, t + \delta_2\, t^2 + u_{it} + v_{it} \qquad (5)$$

We suggest a second order polynomial to account for possible nonlinear nature of the market structure change.

In the second extension, we add 5 country dummies cn_i to the baseline model to capture differences across countries and to provide a ranking of countries in terms of concentration levels. Hence the model becomes:

$$\tilde{C}_{kit} = \beta_0 + \frac{\beta_1}{\ln(S_{it}/\sigma_i)} + \sum_{i}^{5} \varphi_i\, cn_i + u_{it} + v_{it} \qquad (6)$$

There is a concern that including dummy variables may lead to the so-called incidental parameters problem in the context of a non-linear model when estimating without transformation (see Greene, 2005). It was however found that the incidental parameters problem does not cause considerable bias when time dimension is large. This is exactly the

case for our data, which gives us confidence about consistent estimation. Besides, we have also estimated the model in (5) using both true random effect (Greene, 2005) and generalised true random effects (Badunenko and Kumbhakar, 2017) and the results are consistent with what we obtain for the simpler stochastic frontier model.[6] In the third extension, we add both time trend and country dummies to check for the robustness of our model. We then estimate the following:

$$\tilde{C}_{kit} = \beta_0 + \frac{\beta_1}{\ln(S_{it}/\sigma_i)} + \delta_1\, t + \delta_2\, t^2 + \sum_i^5 \varphi_i\, cn_i + u_{it} + v_{it} \qquad (7)$$

The last model is the least restrictive and will provide evidence if the market structure is different across countries and time.

6. Empirical Findings

6.1. Market Structure

Error! Reference source not found. shows market structure indicators of the GCC banking markets over the period 2000–2017. Overall, the GCC banks operate under concentrated market conditions. The banking markets of Bahrain, Kuwait, and Oman are concentrated, where two banks dominate over 50% of the banking market. The level of concentration is higher in Qatar, as only one bank represents half of the market. Yet the Saudi and Emirati banking markets exhibit unconcentrated conditions, considering an HHI of less than 1,500. Throughout, the Qatari banking market is the most concentrated; Bahrain, Oman, and Kuwait are moderately concentrated; and Saudi Arabia and the UAE are the least concentrated markets in the region. The ranking of concentration levels may correspond to the fact that Saudi Arabia

[6] The full set of results are available from authors upon request.

and the UAE present the largest markets in the region, while Qatar, Bahrain, and Oman are the smallest ones.

Moreover, the time series evolution of CR_1 and CR_3 as presented in **Error! Reference source not found.** shows that the share of the first and the three largest banks remains stable over the sample period, indicating that dominant banks maintained the same market share over time. The CR_1 ranges between 20% and 60%, while the CR_3 ranges between 40% and 80%. These measures suggest that the banking markets of the Gulf region are operating under concentrated conditions in general, as the share of the three largest banks dominates up to 80% of the banking markets. In terms of concentration ranking, Qatar and Bahrain seem to be the most concentrated banking markets, while Saudi Arabia and UAE are the least concentrated ones.

6.2. Size–structure relationship

This section provides an analysis of the relationship between market concentration and market size in the Gulf region over the 2000–2017 period. We first follow Sutton's theory and assume that all countries are the same market (Baseline Model). Then, we explore several extensions to account for both time and country heterogeneity on this relationship.

6.2.1. Baseline Model

This section presents the results of the baseline model as in Sutton's theoretical framework. **Error! Reference source not found.** depicts this relationship based on two measures of concentration, CR1 and CR3, as well as two measures of market size, total banking assets and population. The measures of market size are presented in natural log to facilitate the

appreciation of figures. The figures illustrate the observed concentration in markets with as small as 22 constant billion USD and as large as 1,203 constant billion USD in terms of total banking assets. In terms of population, the market size ranges from as few as 442,000 people to as many as 24 million. The figure also demonstrates a range of values for both CR1 and CR3. Though CR1 is 34% on average, it ranges between 15% in the UAE and 56% in Bahrain. The average value of CR3 is 63%, and again it ranges from 40% to 84%.

Noticeably, there seems to be a lower bound to concentration amongst all market sizes. For both CR1 and CR3, the concentration levels peak at the small markets, decrease for a range, but hit a lower limit beyond which concentration levels do not fall. This pattern is robust for the two alternative measures of market size (total banking assets, population). In all cases, the concentration levels remain strictly bounded below, and apparently concentration does not fragment in the GCC banking industry.

Estimates of the lower bound using the time-variant SF model are illustrated in **Error! Reference source not found.**, whilst the parameter estimates and their standard errors are presented in **Error! Reference source not found.**. Here, the concentration measures are presented in their logit transformation, whereby CR1 becomes ln (CR1/1-CR1) and CR3 becomes ln (CR3/1-CR3). Also, the measures of market size are scaled by the set-up costs. The solid lines in **Error! Reference source not found.** present the estimates of the lower bounds using the SF model of Eq. 4. Note that there are observations below the lower bound, which is a result of making the Sutton model stochastic. Focusing on CR1, the estimates show remarkably flat lower bounds. Even though the lower bounds of CR3 are slightly decreasing for larger market sizes, they remain asymptotic to positive levels. This is shown by the strictly positive values of the limiting level of concentration as market size grows to infinity (C_1^{∞} and C_3^{∞}). C_1^{∞} is 19% for the first measure of market size (assets) and 18% for the second

(population), with 95% confidence intervals [16.1; 22.7] and [10.9; 24.9], respectively. Similarly, the limiting levels of CR3 are 28% and 22%, with corresponding 95% confidence intervals [23.1; 33.8] and [17.3; 26.4]. All intervals are bounded well above zero.

These estimates provide convincing evidence that concentration does not fall as market size grows. Therefore, the banking industry of the GCC is a natural oligopoly and remains concentrated around some level regardless of the market size. The finding proposes that the GCC banking industry is of an endogenous sunk-cost type, where quality improvements rely mainly on fixed rather than variable costs. Hence, the equilibrium level of concentration is invariant to variations in market size. This is the central idea of Sutton's theoretical rationale, where market expansion leads the current dominant banks to raise the fixed investment in quality enhancement. In larger markets, thus, there will not be a larger number of banks but instead a better level of service quality that is provided mainly by dominant banks. In the banking context, the quality investments can include building more branches, expanding the ATM network, and advertising more. Similar findings are also reported by Dick (2007) for the US banking industry.

For a robustness check, we also estimate parameters in Eq. 3 and the lower bound using minimum distance implemented as a linear programming problem. More specifically, we minimise the sum of distances from observations to the lower bound where these distances are restricted to be non-negative. Consequently we use the maximum likelihood estimation approach to obtain the standard errors of the parameters, based on the assumption that ε_{it} in Eq. 3 is distributed as a two-parameter Weibull, as originally proposed by Sutton (see Dick (2007) for further details on this two-step procedure). In this case, none of the observations are under the lower bound, but the lower bound is not stochastic. The empirical estimations and the associated lower bounds are shown in Appendix B. The estimations confirm the existence

of a lower bound to concentration that converges to asymptotically positive levels. The limiting levels of concentration of CR1 are 9% for the first measure of market size (assets) and 4% for the second measure (population). These values have 95% confidence intervals between [2.9; 9.5] and [3.8; 5.1], respectively. Likewise, the limiting levels of CR3 are 21% and 12% with 95% confidence intervals [18.2; 22.9] and [11; 13.5], respectively. All limiting levels are strictly positive and bounded well at positive levels. These findings confirm that the GCC banking industry is an endogenous sunk-cost industry where fixed costs play a central role in defining the equilibrium structure.

6.2.2. Extension 1: Cross-time heterogeneity

We introduce a time trend to Sutton's base model to capture any heterogeneity across time in the limiting level of concentration. Because the two measures of concentration yield robust outcomes, we will proceed with CR1 only.

The estimation results of Eq. 5 that incorporate a time trend in the baseline model are presented in **Error! Reference source not found.**, and the associated lower bounds are displayed in **Error! Reference source not found.**. Over the period of 2000–2017, the lower bound is flat for the first measure of market size (assets) and has a slightly decreasing trend for the other measure (population). Yet again, the limiting values of the estimated bounds as market size goes to infinity are strictly positive and bounded well above zero. The mean value of the limiting level of concentration is 17% and bounded between 13.5% and 20.6% for the first measure (assets), whereas it is 6% and bounded between 3.6% and 9% for the second measure (population).

Error! Reference source not found. illustrates a timeline for the limiting level of concentration as market grows to infinity over the sample period. The liming levels of concentration are strictly positive and also bounded at positive levels for all years of our sample. Surprisingly, the limiting levels are not only significantly positive but also exhibit a rising trend over time. These findings confirm those of the baseline model. In other words, the banking industry of the GCC does not fragment regardless of the market size, and the dominant banks maintain their market share even through a long period of time. The limiting level of concentration is eventually increasing across time. This finding suggests that the largest banks may have become more dominant over time. A possible explanation goes back to the central notion of Sutton's theory—that as markets grow larger, the dominant bank will increase the level of fixed-cost investment to reflect additional consumer demand. Thus, dominant banks would have an increased market share over time.

6.2.3. Extension 2: Cross-country heterogeneity

In the second extension, we aim to investigate the differences in the equilibrium levels of concentration across the six Gulf countries by incorporating country dummies in Sutton's model (Eq. 6). **Error! Reference source not found.** presents the estimation results. The estimated coefficients of the country dummies are positive and significant, which confirm the heterogeneity across the GCC countries. Also, the limiting values of concentration vary considerably across different countries. The limiting levels range from 20% in the UAE to 50% in Qatar for the first measure of market size (Assets), whilst they are between 19% and 42% for the other measure (Population). This emphasises the robustness of findings across alternative measures of market size.

Error! Reference source not found. further displays the ranking of the GCC countries according to their limiting levels of concentration. For both measures of market size, the equilibrium levels of concentration are the highest for Qatar, followed by Bahrain and Oman, whereas the limiting levels are the lowest in the UAE and Saudi Arabia. Kuwait, on the other hand, lies in the middle. All limiting levels are strictly positive and bounded at positive levels.

Interestingly, the ranking of countries in terms of their limiting levels of concentration is consistent with the ranking according to the actual concentration levels. Another remarkable finding is that the values of the limiting and actual levels of concentration are astonishingly close. This indicates that the banking markets of the GCC countries are operating under a long-run equilibrium.

The estimates of the lower bounds are shown in **Error! Reference source not found.**. The estimated lower bounds for all countries are outstandingly flat, and this result is robust for both measures of market size. This confirms Sutton's theory for endogenous sunk-cost industries, where the levels of concentration are invariant to changes in market size. As markets grow, the number of banks will not necessarily increase, but instead the dominant banks in the GCC will escalate their fixed investments to improve the quality of services and hence raise the barriers to entry.

6.2.4. Extension 3: Time and country heterogeneity

In the last extension, we propose the inclusion of both a time trend and country dummies in the baseline model of Sutton. At this juncture, we will investigate the heterogeneity in

concentration over time and across countries. **Error! Reference source not found.** presents the estimation results of Eq. 7.

For the first measure of market size (assets), the results substantiate those of extension 2. The limiting levels of concentration peak again for Qatar around 45%, while the UAE exhibits the lowest levels of around 19%. Saudi Arabia ranks second after the UAE with the least limiting levels of concentration in the region, with C_1^{∞} equal to 21%. Likewise, the estimations imply that Kuwait is ranked third, with least limiting levels of concentration of about 28%. For Bahrain and Oman, which exhibit similar market sizes in terms of total banking assets, the limiting levels of concentration are also notably parallel, at around 40%.

The limiting levels of concentration stand between 10% and 21% for the second measure of market size (population). Although the findings are different quantitatively, they are equivalent qualitatively. In terms of ranking, the limiting levels remain at maximum in Qatar, Oman, and Bahrain and exhibit the minimum levels in the UAE, Saudi Arabia, and Kuwait.

The timeline evolution of the limiting levels of concentration over time and countries is set out in **Error! Reference source not found.**. All values are at positive levels with positive 95% confidence intervals for all years and countries. It can be clearly seen that the limiting levels of concentration provide a ranking of GCC countries from lowest to highest as follows: the UAE, Saudi Arabia, Kuwait, Bahrain, Oman, and Qatar. This also corresponds to the size of banking markets in each country. The banking markets are the largest in the UAE and Saudi Arabia and the smallest in Oman and Bahrain. Consequently, it can be proposed that though the limiting levels of concentration will always be restricted to positive levels regardless of size, there can be variation in the levels themselves because of variances in market size.

Moreover, the estimates of the lower bound across time and countries[7] are flat and strengthen the conclusion that the banking industry of the GCC is an endogenous sunk-cost industry where shifting the consumer demand relies heavily on fixed rather than variable costs. Because this is not feasible for all banks in the market, it represents a natural barrier to entry and defines an oligopolistic market structure with a few dominant banks and a number of fringe banks.

These results provide important insights into the market structure of the GCC countries. First, the concentration measures indicate that the GCC banking market is operating under somewhat concentrated conditions except for the UAE and Saudi Arabia. Second, there exists a lower bound to concentration, and the market remains concentrated regardless of size. This implies that the banking industry of the GCC is an endogenous sunk-cost industry. Third, the limiting levels of concentration are positive and bounded well above zero. Fourth, the limiting levels are increasing over time, and so the largest banks become more dominant. Fifth, the banking markets of the GCC can be ranked according to their limiting levels of concentration as Qatar, Oman, and Bahrain with the highest levels, the UAE and Saudi Arabia with the lowest levels, and Kuwait lying somewhat in the middle. Finally, all results are robust for alternative measures of market size and also across countries and time.

7. Discussion

As the literature suggests, market size and the level of market concentration are expected to be inversely related. Expansion of profits with market growth induces new firms to enter the market. This will reduce the market shares of incumbents, and hence concentration levels will

[7] To save space, we do not include the graph which bears similarity to Figure 8, however, it is available upon request.

fall. Sutton (1991) proposed that this negative relationship is valid only for certain groups of industries and that the relationship breaks down for industries in which enhancing consumers' willingness to pay relies predominantly on sunk costs.

The most obvious finding to emerge from the analysis is that there exists a lower bound to concentration that converges asymptotically to positive levels in the GCC banking markets. Additionally, the market structure remains concentrated throughout all market sizes. This general characteristic provides rich findings. The banking industry of the GCC seems to be of an endogenous sunk-cost type, where fixed costs play a fundamental role in determining the equilibrium market structure. A possible explanation for this finding is that as markets expand, dominant banks will raise the provision of quality through incurring additional fixed costs. These quality investments can include building more branches and expanding the ATM network to serve larger markets and reflect greater consumer demand. Also, quality investments could be in terms of technological advances to provide banks with the means to provide innovative online banking services. Notwithstanding this, banks may augment the level of advertising to attract consumers. Such continuous escalation in fixed costs is not always feasible for all banks, thereby making potential entry harder. In such industries, thus, the number of entrants may be greatly invariant to changes in market size.

Over the sample period, another interesting finding is that the limiting levels of concentration eventually trend upward. This could be because dominant banks become even more dominant over time as their attempts to enhance service quality through greater sunk costs further amplify their market share and consequently the level of market concentration. This raises a significant policy implication for central banks. While more control is recommended for larger banks to avoid systemic risk, the growth of market share could be attributed to providing a higher quality of services and not necessarily engaging in risky portfolios.

Across individual countries, the empirical results depict some heterogeneity in the limiting levels of concentration. The limiting levels are the greatest for Qatar, Bahrain, and Oman, while they are the least for the UAE and Saudi Arabia. Kuwait, on the other hand, lies somewhat in the middle. This heterogeneity can be attributed to several explanations. First, the level of set-up costs in the banking market of the country can denote a huge barrier to entry. Excessive set-up costs discourage entrance to the industry and raise the concentration levels. Throughout the GCC countries, the level of set-up costs is at a minimum in Saudi Arabia and the UAE and at a maximum in Qatar.[8] Moreover, the ranking of GCC countries in terms of the level of set-up is precisely identical to their rankings in terms of the limiting levels of concentration. Another possible explanation for this heterogeneity is the market size. At the country level, larger markets exhibit the least limiting levels of concentration. With these findings we propose that although the banking industry of the GCC is of an endogenous sunk-cost type that remains concentrated regardless of size, the values of limiting levels of concentration can depict some variations across individual countries due to the level of set-up costs and market size. In other words, the market structure of the banking industry in the Gulf region remains concentrated; however, there is heterogeneity in the levels of concentration across countries due to set-up costs and market size. For instance, the UAE faces low levels of set-up costs and represents the largest market in the region. At the same time, the UAE exhibits the least levels of equilibrium structure.

These findings offer interesting implications to antitrust authorities who focus mainly on market concentration measures when approving M&A requests. Among all, the quality of services provided by a bank should also be taken into account when assessing any merger or acquisition request. So, a relevant question might be whether the new bank would deliver

[8] This information is according to the banking regulations of the central bank of each country.

greater quality of services that in turn would contribute towards greater consumer welfare. In this context, an important aspect to look at for example will be whether the merged bank will broaden the ATM or branch network, or make any other quality improvements. This leads to an unforeseen conclusion that market concentration may not always be harmful for consumer welfare.

8. Conclusion

This study investigates the market structure and the size–structure relationship in the GCC banking industry over the period 2000–2017. The basis of the analysis comes from the model of Sutton (1991), which we estimate using a stochastic frontier approach. To test the hypotheses that the prediction of the Sutton model may be different for different time periods of segments of the whole market, we provide several extensions in the panel data context.

The main results can be summarised as follows. Based on the concentration measures of individual GCC countries, we establish the stylised fact that most of the GCC banking markets are highly concentrated except for the UAE and Saudi Arabia, where the levels of concentration are quite low. Furthermore, in this region, the countries are ranked similarly according to both concentration levels and set-up costs.

The empirical analysis of the GCC banking single market shows that there exists a lower bound to concentration, and the market remains concentrated regardless of market size. That being the case, the banking industry of the GCC is an endogenous sunk-cost industry, which means that sunk costs (mainly the set-up costs) in this market represent a substantial barrier for the potential newcomers.

This result holds both across time and individual GCC countries, as the limiting levels of concentration in the GCC banking markets remain positive and bounded well above zero. This

result emphasizes that the market size is not related to the market structure in the GCC banking industry. Thus, sunk costs are the determinant of the long-run equilibrium structure in the GCC banking industry. For this reason, the GCC banking market has a naturally concentrated structure.

Furthermore, the limiting levels of concentration vary across countries and the variation is in line with the actual level of concentration. These findings lead to the proposition, that although the banking industry of the GCC is of an endogenous sunk-cost type that remains concentrated regardless of size, the values of limiting levels of concentration exhibit some variations across individual countries, due to the level of set-up costs and market size. Finally, the limiting levels are increasing over time, and so the largest banks become more dominant. This suggests that the incumbent banks have encountered higher sunk costs in order to improve the banking services, however, all results taken together show the incumbents can benefit from the large market share that they have.

Coupled with the theoretical predictions, our results suggest that a concentrated banking market and the existence of incumbent banks may not always be detrimental to consumer welfare. Incumbents may enjoy a higher market share through expanding the sunk costs in improving the quality of services provided to consumers and hence enhancing consumer welfare.

References

Abuzayed, B., Al-Fayoumi, N., Molyneux, P., 2018. Diversification and bank stability in the GCC. J. Int. Financ. Mark. Institutions Money 57, 17–43. https://doi.org/10.1016/j.intfin.2018.04.005

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. J. Econom. 6, 21–37. https://doi.org/https://doi.org/10.1016/0304-4076(77)90052-5

Al-Muharrami, S., Matthews, K., Khabari, Y., 2006. Market structure and competitive conditions in the Arab GCC banking system. J. Bank. Financ. 30, 3487–3501. https://doi.org/10.1016/j.jbankfin.2006.01.006

Ashraf, D., Ramady, M., Albinali, K., 2016. Financial fragility of banks, ownership structure and income diversification: Empirical evidence from the GCC region. Res. Int. Bus. Financ. 38, 56–68. https://doi.org/10.1016/J.RIBAF.2016.03.010

Badunenko, O., Kumbhakar, S., 2017. Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? Eur. J. Oper. Res. 260, 789–803. https://doi.org/10.1016/j.ejor.2017.01.025

Bain, J.S., 1956. Barriers to new competition: Their character and consequences in manufacturing industries. Harvard University Press., Cambridge, Mass.

Basu, R., Prasad, A., Rodriguez, S., 2018. Monetary operations and Islamic banking in the GCC: Challenges and options. J. Gov. Regul. 7, 49–63.

https://doi.org/10.22495/jgr_v7_i1_p4

Claessens, S., van Horen, N., 2014. Foreign banks: Trends and impact. J. Money, Credit Bank. 46, 295–326. https://doi.org/10.1111/jmcb.12092

Clerides, S., Delis, M.D., Kokas, S., 2015. A new data set on competition in national banking markets. Financ. Mark. Institutions Instruments 24, 267–311. https://doi.org/10.1111/fmii.12030

Dick, A.A., 2007. Market size, service quality, and competition in banking. J. Money, Credit Bank. 39, 49–81. https://doi.org/10.1111/j.0022-2879.2007.00003.x

Dong, Y., Firth, M., Hou, W., Yang, W., 2016. Evaluating the performance of Chinese commercial banks: A comparative analysis of different types of banks. Eur. J. Oper. Res. 252, 280–295. https://doi.org/10.1016/j.ejor.2015.12.038

Ellickson, P.B., 2007. Does Sutton apply to supermarkets? RAND J. Econ. 38, 43–59. https://doi.org/10.1111/j.1756-2171.2007.tb00043.x

Greene, W., 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J. Econom. 126, 269–303. https://doi.org/10.1016/j.jeconom.2004.05.003

Greene, W., 1990. A gamma-distributed stochastic frontier model. J. Econom. 46, 141–163. https://doi.org/https://doi.org/10.1016/0304-4076(90)90052-U

Herfindahl, O., 1950. Concentration in the steel industry. (Unpublished doctoral thesis). Columbia University, New York, NY.

Hirschman, A., 1964. The paternity of an index. Am. Econ. Rev. 54, 761–762.

Kumbhakar, S.C., Lovell, C.A.K., 2003. Stochastic frontier analysis. Cambridge University Press., Cambridge.

Mason, E.S., 1939. Price and production policies of large-scale enterprise. Am. Econ. Rev. 29, 61–74.

Ondrich, J., Ruggiero, J., 2001. Efficiency measurement in the stochastic frontier model. Eur. J. Oper. Res. 129, 434–442. https://doi.org/https://doi.org/10.1016/S0377-2217(99)00429-4

Park, B.U., Sickles, R.C., Simar, L., 1998. Stochastic panel frontiers: A semiparametric approach. J. Econom. 84, 273–301.

Ritter, C., Simar, L., 1997. Pitfalls of normal-gamma stochastic frontier models. J. Product. Anal. 8, 167–182. https://doi.org/10.1023/A:1007751524050

Shaked, A., Sutton, J., 1987. Product differentiation and industrial structure. J. Ind. Econ. 36, 131-146.

Shaked, A., Sutton, J., 1983. Natural oligopolies. Econometrica 51, 1469–1483.

Shaked, A., Sutton, J., 1982. Relaxing price competition through product differentiation. Rev. Econ. Stud. 49, 3–13.

Sutton, J., 1991. Sunk costs and market structure: Price competition, advertising, and the evolution of concentration. MIT Press., Cambridge, Mass.

Tabacco, G.A., 2013. A new way to assess banking competition. Econ. Lett. 121, 167–169. https://doi.org/10.1016/j.econlet.2013.07.028

Tsionas, M.G., 2017. Microfoundations for stochastic frontiers. Eur. J. Oper. Res. 258, 1165–1170. https://doi.org/10.1016/j.ejor.2016.09.033

U.S. Department of Justice, The Federal Trade Commission, 2010. Horizontal Merger Guidelines. Retrieved from: https://www.ftc.gov/sites/default/files/attachments/merger-review/100819hmg.pdf.

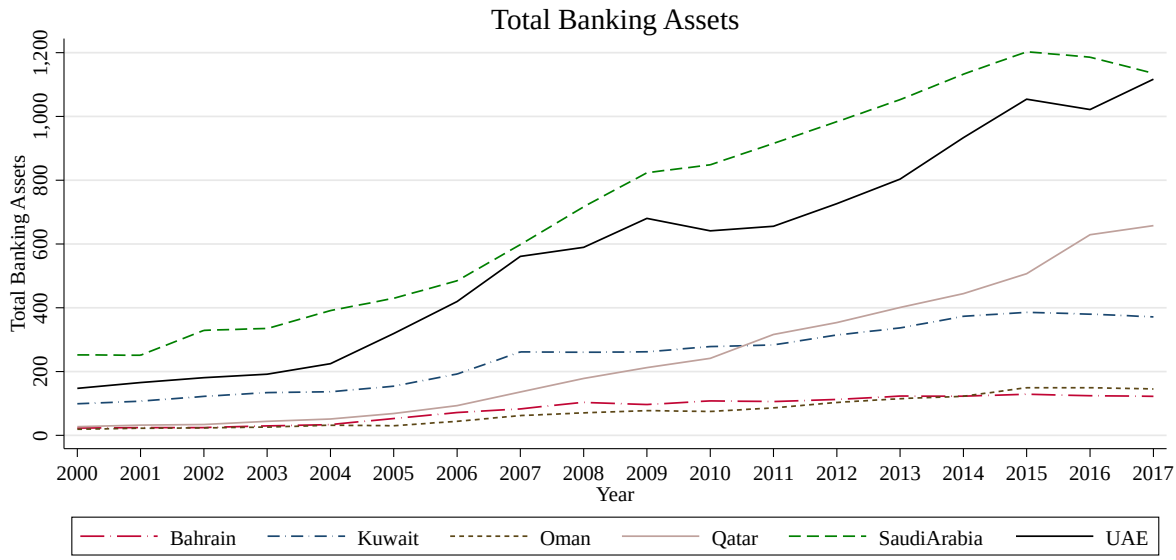

Figure 1. Time series evolution of GCC banking markets. This graph shows a timeline of the total banking assets of all national–retail banks in the GCC between 2000–2017. The total banking assets are reported in constant billion US dollars.

Table 1
Summary statistics. This table provides mean values of population, banking assets, revenues, and set-up costs in the GCC countries over the period 2000–2017. 'Total Banking Assets' is the sum of national retail banks' assets in each year. 'Total Banking Revenues' is the sum of national retail banks' revenues in each year. 'Set-up costs' are the sum of the minimum paid-up capital required and the licensing fees. 'Number of Banks' is the number of banks that have been operating in each country during the sample period. The sample includes the entire population of banks. All values are reported in million $.

	Population	*Total Banking Assets*	*Total Banking Revenues*	*Set-up Costs*	*Number of Banks*
Bahrain	1,091,652	47,003	2,466	265.958	13
Kuwait	2,911,110	144,814	7,887	247.239	10
Oman	3,100,148	35,307	1,892	260.091	9
Qatar	1,528,000	146,920	7,314	274.739	10
Saudi Arabia	26,600,000	343,046	17,259	0.667	12
UAE	6,666,066	314,156	16,305	10.892	25

Table 2

Banking market structure and size of the GCC. This table provides indicators of concentration and market size of the GCC banking sectors over the period 2000–2017. The values represent the sample means. The CR_k ratio is measured by summing the market shares as measured by total assets of top k banks in the market. Correspondingly, CR_1, CR_2, CR_3, and CR_5 represent the market share of the top one, two, three, and five banks, respectively. The ratio values range between zero in a highly competitive market and unity in a monopolistic market. HHI is the sum of squared market shares of banks in a particular banking market. According to the 2010 Horizontal Merger Guidelines, the markets are classified into three types: unconcentrated markets (HHI < 1500), moderately concentrated markets (1500 < HHI < 2500), and highly concentrated markets (HHI > 2500). Assets/Set-up costs is a ratio of deflated total banking assets of a country to deflated set-up costs of a bank; Population/Set-up costs is a ratio of the population of the country divided by deflated set-up costs of a bank.

	CR_1	CR_2	CR_3	CR_5	HHI	$\frac{\text{Assets}}{\text{Setup costs}}$	$\frac{\text{Population}}{\text{Setup costs}}$
Bahrain	44%	59%	74%	87%	2,478	177	2,305
Kuwait	30%	52%	64%	81%	1,886	586	6,755
Oman	42%	59%	73%	91%	2,553	136	5,449
Qatar	48%	61%	72%	87%	2,713	535	3,198
Saudi Arabia	21%	37%	49%	69%	1,282	514,568	18,500,000
UAE	20%	35%	47%	66%	997	28,843	323,121

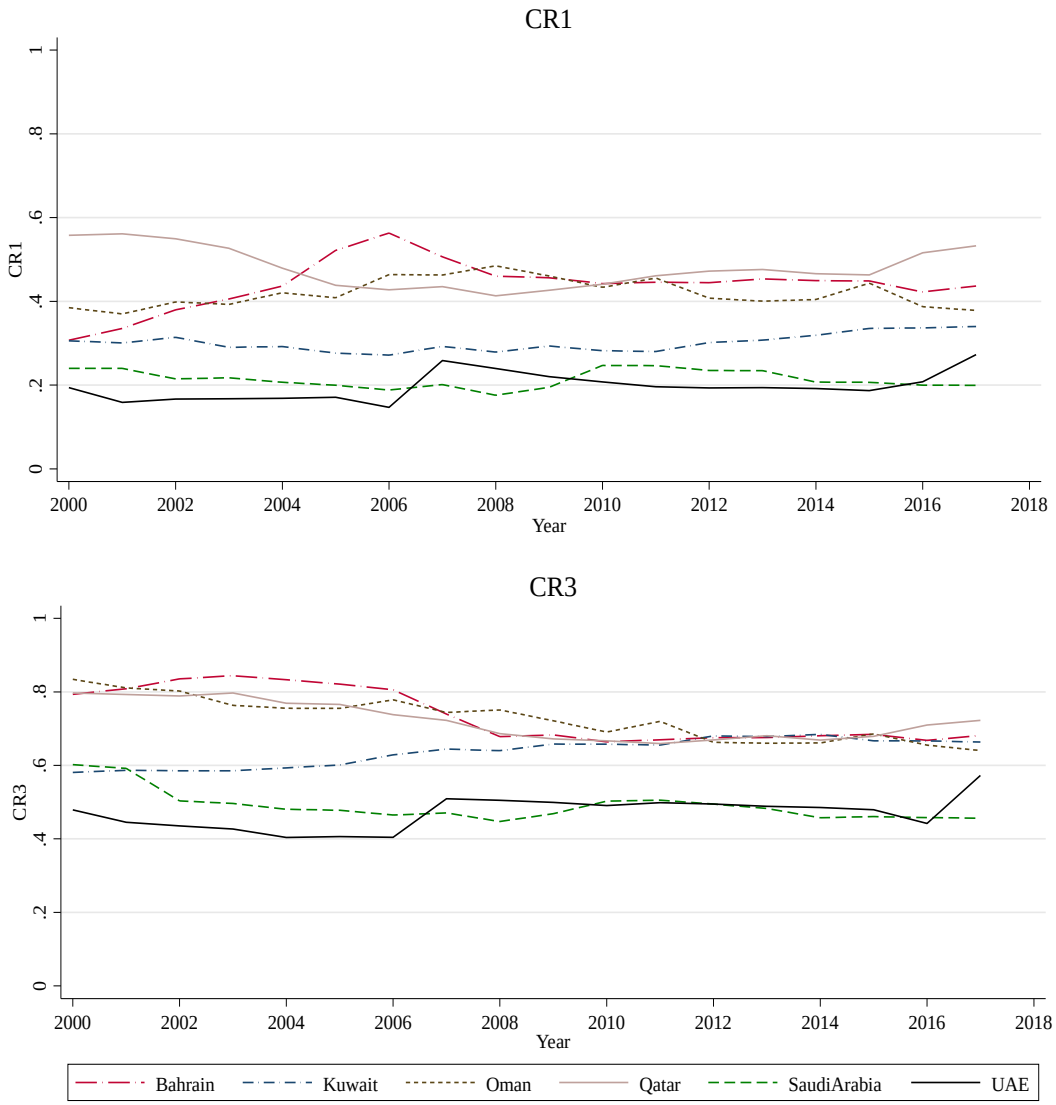

Figure 2. Time series evolution of concentration in the GCC banking markets. This graph shows the concentration levels in the GCC banking markets over the period 2000 to 2017. The upper panel uses CR_1 as a measure of concentration, while the lower panel uses CR_3 as a measure of concentration. CR_1 and CR_3 represent the market share of the top one and three banks, respectively. Market share is measured by the total banking assets.

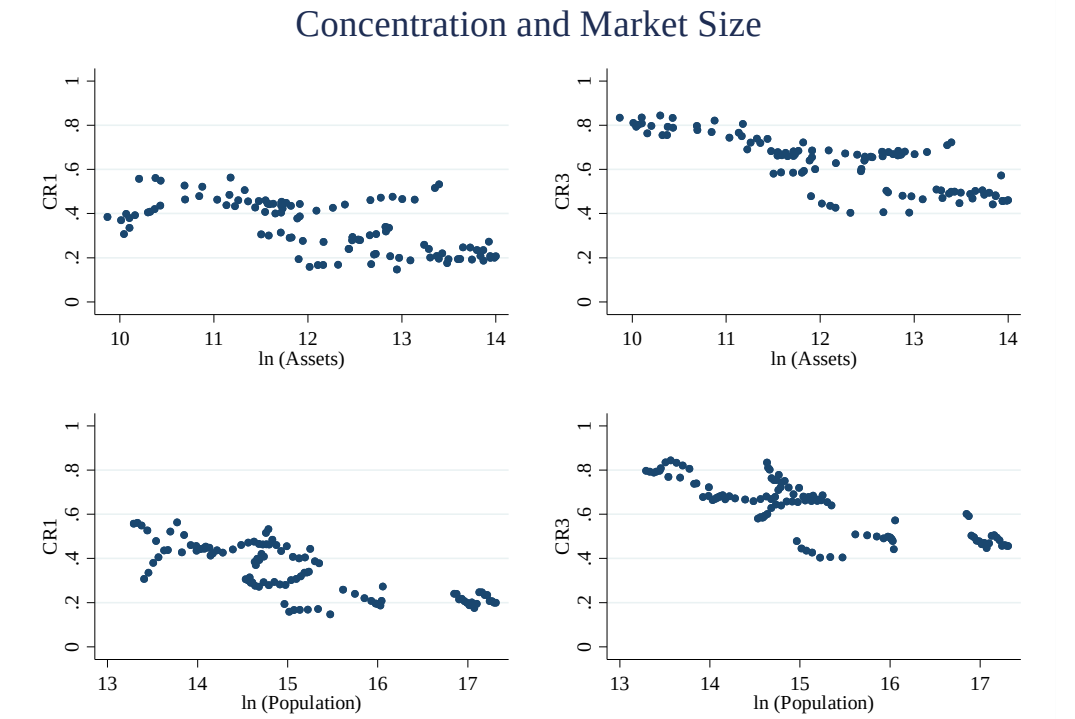

Figure 3. Concentration and market size in the GCC banking markets. This figure contains scatter plots of the size–structure relationship of the GCC banking market over the 2000–2017 period. In the upper panels, market size is measured by ln (Assets), which corresponds to the natural log of deflated total banking assets, whereas the lower panels use ln (Population) as another measure of market size in accordance with Sutton's theoretical model. In the left panels, market structure is measured by a one-bank concentration ratio (CR1), while the right panel presents the share of the three largest banks (CR3).

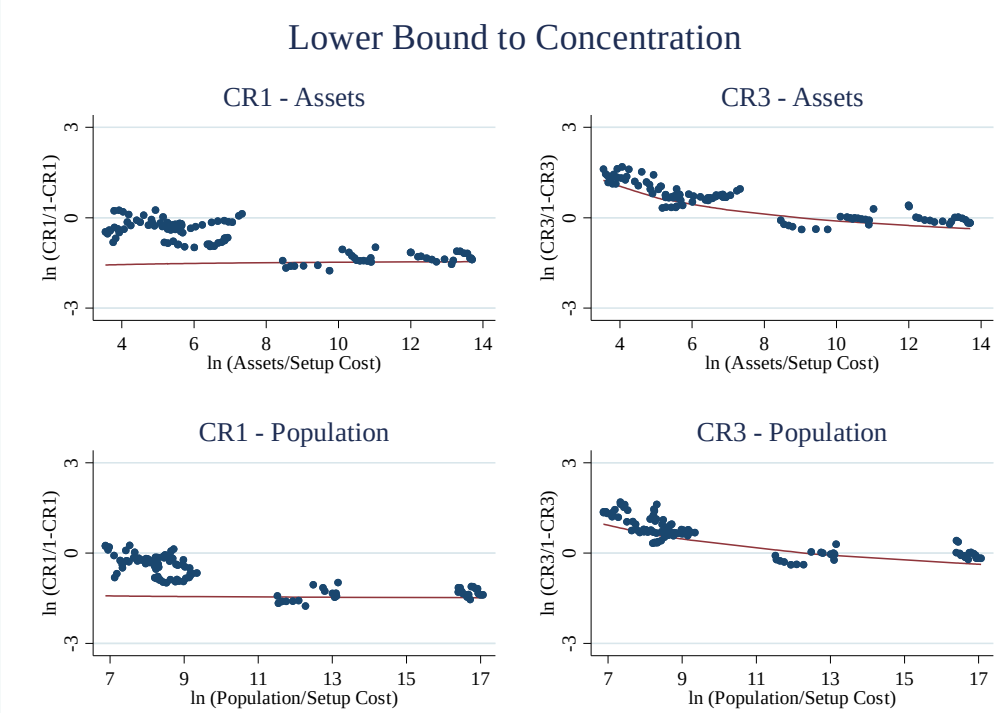

Figure 4. Lower bound to concentration of the baseline model. The solid line represents the stochastic frontier estimates. The upper panel uses the scaled market size, measured as the natural log of assets to set-up costs ratio, whilst the lower panel presents the scaled market size measured as the natural log of population to set-up costs ratio. Concentration measures are presented in their logit transformation. In the left panels, the concentration ratio is ln (CR1/1-CR1), whereas concentration is ln (CR3/1-CR3) in the right panels.

Table 3

Baseline model estimation results. The table provides lower bound estimates of the baseline model for GCC banks over the period 2000–2017. Standard errors are in parenthesis. The dependent variables ($\tilde{C}_1$ and $\tilde{C}_3$) are the logit transformation of CR_1 and CR_3, respectively. The independent variables are the two measures of market size (assets and population). Assets is the natural logarithm of total banking assets over set-up costs; Population is the natural logarithm of population over set-up costs. β_0 and β_1 are the estimated parameters and correspond to the intercept and slope, respectively; μ is the estimated inefficiency; η shows whether inefficiency changes over time; σ^2 is the variance of the composed disturbance term; and γ is a ratio of the inefficiency variance to the disturbance variance. C^{∞} is the limiting level of concentration reached at infinite market size.

	Assets		**Population**	
	$\tilde{C}_1$	$\tilde{C}_3$	$\tilde{C}_1$	$\tilde{C}_3$
β_0	-1.424***	-0.922***	-1.523***	-1.272***
	(0.108)	(0.134)	(0.243)	(0.136)
β_1	-0.526	7.812***	0.695	15.38***
	(0.612)	(1.156)	(2.934)	(1.400)
μ	0.451			
	(0.871)			
η		-0.0291		0.0959***
		(0.0348)		(0.0187)
σ^2	0.614	0.170*	0.844	0.0473***
	(0.733)	(0.0957)	(0.517)	(0.0117)
γ	0.949***	0.769***	0.963***	0.263
	(0.0611)	(0.135)	(0.0236)	(0.167)
Ln σ^2	-0.488	-1.775***	-0.170	-3.050***
	(1.194)	(0.564)	(0.613)	(0.247)
Lgt γ	2.928**	1.202	3.246***	-1.033
	(1.268)	(0.761)	(0.656)	(0.863)
C_1^{∞} ; C_3^{∞}	19%	28%	18%	22%
95% Confidence Interval	[16.1; 22.7]	[23.1; 33.8]	[10.9; 24.9]	[17.3; 26.4]
Observations	108	108	108	108

Table 4

Extension 1 estimation results. The table provides estimation results of Extension 1, where we add a time trend to the baseline model. Standard errors are in parenthesis. The dependent variable is $\tilde{C}_1$ which is the logit transformation of CR1. The independent variables are the two measures of market size, assets and population, as well as the time trend. Assets is the natural logarithm of deflated total banking assets over deflated set-up costs; Population is the natural logarithm of population over deflated set-up costs; t is the time trend; t^2 is the square term of the time trend; σ^2 is the variance of the composed disturbance term; and γ is a ratio of the inefficiency variance to the disturbance variance. C_1^{∞} is the limiting level of concentration reached at infinite market size.

	Assets $\tilde{C}_1$	**Population** $\tilde{C}_1$
β_0	-1.734***	-2.938***
	(0.197)	(0.267)
β_1	1.302	15.76***
	(1.131)	(2.549)
t	0.0200	0.0289**
	(0.0171)	(0.0140)
t^2	-0.000445	-0.000418
	(0.000754)	(0.000695)
σ^2	0.712*	0.205*
	(0.431)	(0.114)
γ	0.957***	0.855***
	(0.0267)	(0.0829)
Ln σ^2	-0.340	-1.583***
	(0.605)	(0.557)
Lgt γ	3.105***	1.774***
	(0.651)	(0.668)
C_1^{∞}	17%	6%
95% Confidence Interval	[13.5; 20.6]	[3.6; 9]
Observations	108	108

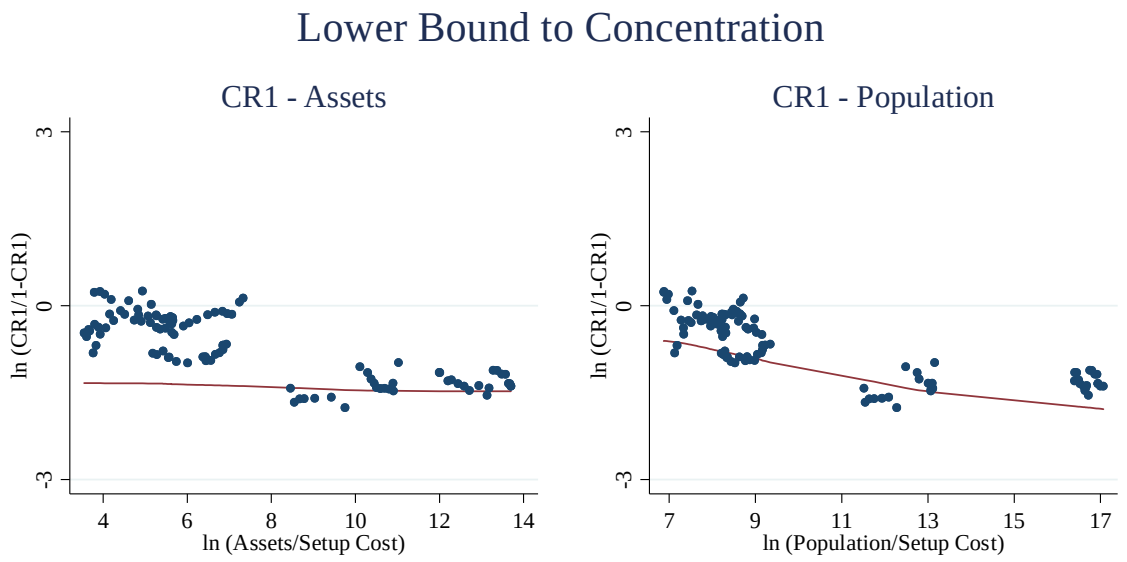

Figure 5. Lower bound to concentration of Extension 1. Measure of concentration is presented in the logit transformation of CR1. The left panel uses 'Assets' as a proxy for market size, measured by the natural log of assets to set-up costs ratio, whilst the right panel presents 'Population' as an alternative proxy of market size, measured by the natural log of population to set-up costs ratio. The solid line represents the stochastic frontier estimates. The scatters are plots of ln (CR1/1-CR1) against market size.

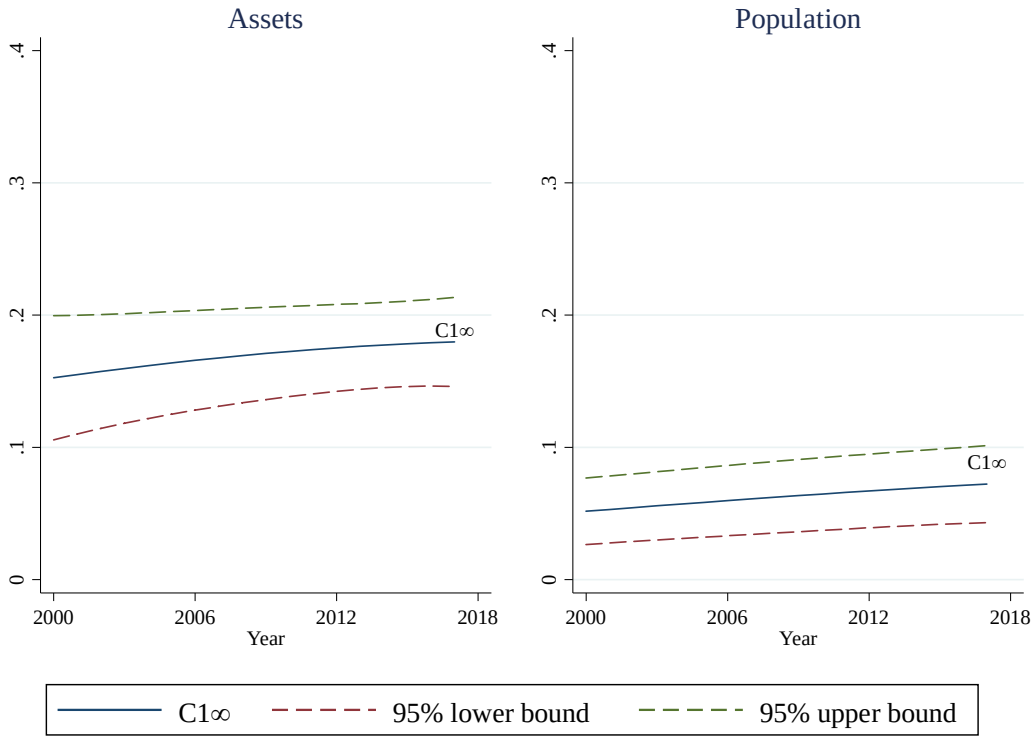

Figure 6. The limiting level of concentration across time. This graph provides a timeline of the limiting level of concentration for the GCC banking industry over the period 2000–2017. The dashed lines are the associated 95% confidence intervals. The left panel uses 'Assets' as a proxy for market size, measured by the natural log of assets to set-up costs ratio, whilst the right panel presents 'Population' as an alternative proxy of market size, measured by the natural log of population to set-up costs ratio.

Table 5
Extension 2 estimation results. The table provides estimation results of Extension 2, where we add country dummies to the baseline model. Standard errors are in parenthesis. The dependent variable is $\tilde{C}_1$, which is the logit transformation of CR1. The independent variables are the two measures of market size, assets and population, as well as country dummies. Assets is the natural logarithm of deflated total banking assets over deflated set-up costs; Population is the natural logarithm of population over deflated set-up costs; Bahrain, Kuwait, Oman, Qatar, and Saudi Arabia are country dummies; σ^2 is the variance of the composed disturbance term; and γ is a ratio of the inefficiency variance to the disturbance variance. C_1^{∞} is the mean value of the limiting level of concentration that is reached at infinite market size.

	Assets $\tilde{C}_1$	**Population** $\tilde{C}_1$
β_0	-1.390***	-1.575***
	(0.0806)	(0.337)
β_1	-0.436	1.629
	(0.623)	(3.907)
Bahrain	1.227***	1.107***
	(0.0843)	(0.194)
Kuwait	0.606***	0.524***
	(0.0682)	(0.148)
Oman	1.158***	1.053***
	(0.0918)	(0.143)
Qatar	1.381***	1.268***
	(0.0765)	(0.193)
Saudi Arabia	0.114*	0.165
	(0.0610)	(0.111)
η	-1.182	-0.639
	(1.575)	(1.050)
σ^2	0.0486*	0.0555*
	(0.0274)	(0.0324)
γ	0.416	0.490
	(0.336)	(0.305)
Ln σ^2	-3.023***	-2.892***
	(0.563)	(0.584)
Lgt γ	-0.337	-0.0389
	(1.381)	(1.221)
C_1^{∞}:		
Bahrain	46%	38%
Kuwait	31%	26%
Oman	44%	37%
Qatar	50%	42%
Saudi Arabia	22%	20%
UAE	20%	17%
Observations	108	108

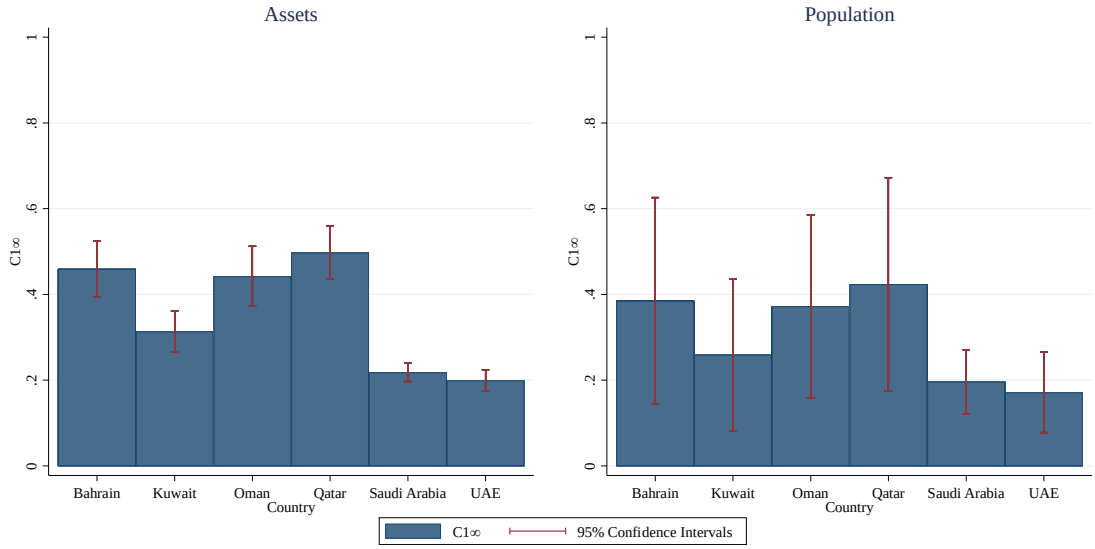

Figure 7. The limiting level of concentration across countries. This graph provides the mean values of the limiting level of concentration for the GCC banking industry over the period 2000–2017. The bars represent the mean values of the limiting levels of concentration. The lines represent the associated 95% confidence intervals. The left panel uses 'Assets' as a proxy for market size, measured by the natural log of assets to set-up costs ratio, whilst the right panel presents 'Population' as an alternative proxy of market size, measured by the natural log of population to set-up costs ratio.

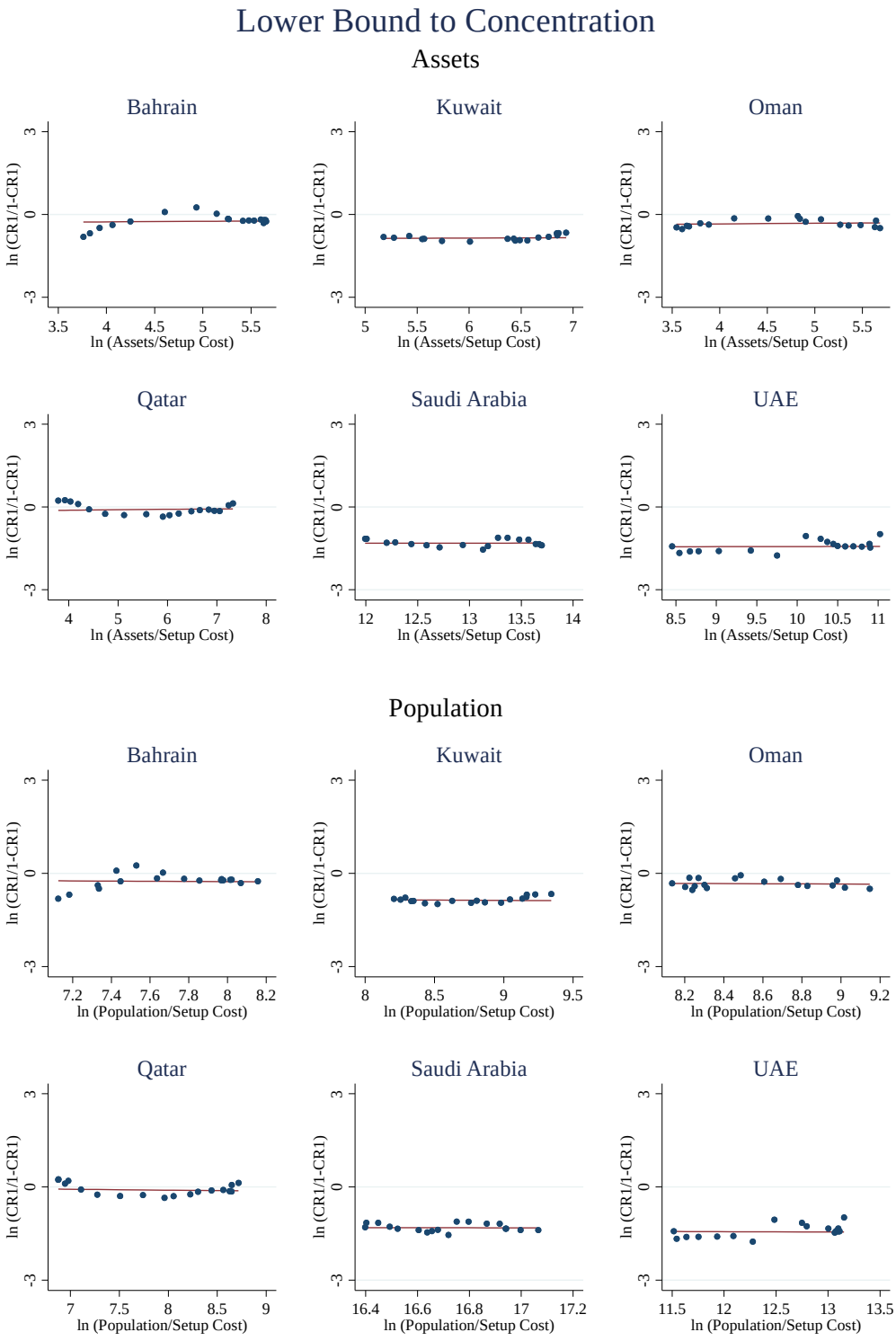

Figure 8. Lower bound to concentration of Extension 2. Measure of concentration is presented in the logit transformation of CR1. The upper panel uses 'Assets' as a proxy for market size, measured by the natural log of assets to set-up costs ratio, whilst the lower panel presents 'Population' as an alternative proxy of market size, measured by the natural log of population to set-up costs ratio. The solid lines represent the SF estimates. The scatters are plots of ln (CR1/1-CR1) against market size.

Table 6
Extension 3 estimation results. The table provides estimation results of Extension 3, where we add both a time trend and country dummies to the baseline model. Standard errors are in parenthesis. The dependent variable is $\tilde{C}_1$ which is the logit transformation of CR1. The independent variables are the two measures of market size, assets and population, a time trend, and country dummies. Assets is the natural logarithm of deflated total banking assets over deflated set-up costs; Population is the natural logarithm of population over deflated set-up costs; Bahrain, Kuwait, Oman, Qatar, and Saudi Arabia are country dummies; σ^2 is the variance of the composed disturbance term; and γ is a ratio of the inefficiency variance to the disturbance variance. C_1^{∞} is the mean value of the limiting level of concentration that is reached at infinite market size.

	Assets	**Population**
	$\tilde{C}_1$	$\tilde{C}_1$
β_0	-1.586***	-2.359***
	(0.181)	(0.432)
β_1	0.657	9.715**
	(1.065)	-4.732
t	0.0142	0.0210
	(0.0164)	(0.0145)
t^2	-0.000288	-0.000307
	(0.000730)	(0.000676)
Bahrain	1.104***	0.685***
	(0.124)	(0.244)
Kuwait	0.534***	0.240
	(0.0862)	(0.172)
Oman	1.015***	0.734***
	(0.140)	(0.184)
Qatar	1.284***	0.871***
	(0.105)	(0.235)
Saudi Arabia	0.130**	0.311***
	(0.0621)	(0.111)
σ^2	0.0288***	0.0278***
	-0.00392	-0.00379
γ	0.00	0.00
	0.00000454	0.00000779
Ln σ^2	-3.547***	-3.581***
	(0.136)	(0.136)
Lgt γ	-20.38	-19.12
	(3231.5)	(1560.5)
C_1^{∞}:		
Bahrain	41%	18%
Kuwait	28%	12%
Oman	39%	19%
Qatar	45%	21%
Saudi Arabia	21%	13%
UAE	19%	10%
Observations	108	108

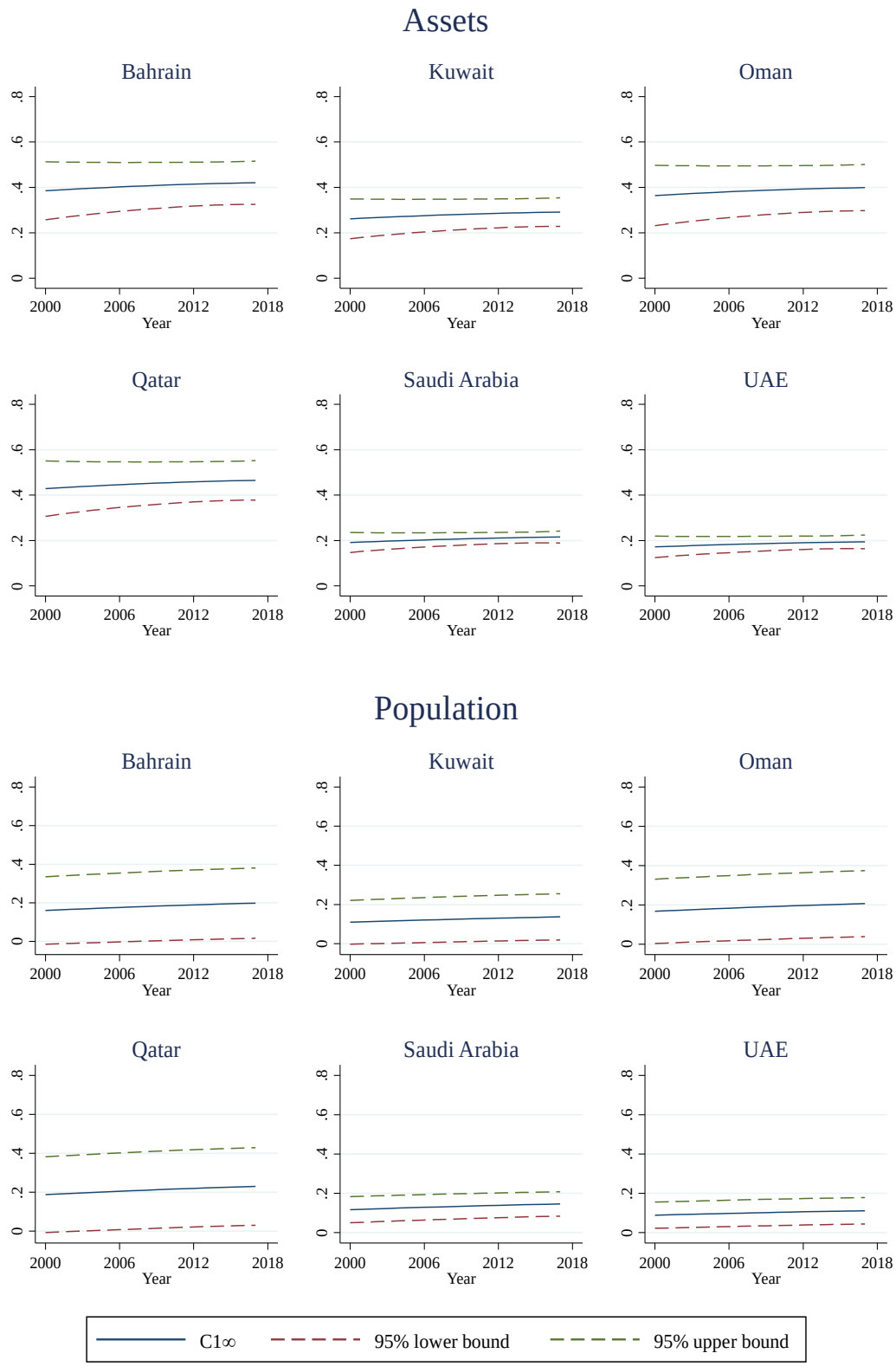

Figure 9. The limiting level of concentration across time and countries. This graph provides a timeline of the limiting level of concentration for the GCC banking industry over the period 2000–2017. The dashed lines represent the associated 95% confidence intervals. The top panel uses 'Assets' as a proxy for market size, measured by the natural log of assets to set-up costs ratio, whilst the lower panel presents 'Population' as an alternative proxy of market size, measured by the natural log of population to set-up costs ratio.

Appendices

Appendix A: List of GCC Banks

	Bank Name	Data availability from	Establish Date
	Bahrain		
1	National Bank of Bahrain	2000	1957
2	Bank of Bahrain & Kuwait	2000	1971
3	Bahrain Islamic Bank	2000	1978
4	Shamil Bank	2000-09	1982
5	Bahraini Saudi Bank	2000-08	1983
6	Al Baraka Islamic Bank	2007	1984
7	Ahli United Bank	2000	2000
8	Kuwait Finance House	2002	2002
9	Khaleeji Commercial Bank	2005	2003
10	BMI Bank	2005-13	2004
11	Future Bank	2005	2004
12	Al-Salam Bank	2006	2005
13	Ithmaar Bank	2010	2010
	Kuwait		
1	National Bank of Kuwait	2000	1952
2	Commercial Bank of Kuwait	2000	1960
3	Gulf Bank	2000	1960
4	Al Ahli Bank of Kuwait	2000	1967
5	Al Ahli United Bank	2000	1971
6	Kuwait International Bank	2000	1973
7	Burgan Bank	2000	1977
8	Kuwait Finance House	2000	1977
9	Boubyan Bank	2005	2004
10	Warba Bank	2011	2010
	Oman		
1	HSBC Bank Oman	2000	1948
2	National Bank of Oman	2000	1973
3	Oman Arab Bank	2000	1973
4	Bank Muscat	2000	1982
5	Bank Dhofar	2000	1990
6	Ahli Bank	2000	1998
7	Bank Sohar	2007	2007
8	Bank Nizwa	2012	2012
9	Al Izz Islamic Bank	2013	2012
	Qatar		
1	Qatar National Bank	2000	1965
2	Commercial Bank	2000	1975
3	Doha Bank	2000	1979

4	Qatar Islamic Bank	2000	1983
5	Al Ahli Bank	2000	1984
6	Qatar International Islamic Bank	2000	1991
7	International Bank of Qatar	2001	2000
8	Masraf Al Rayan	2006	2006
9	Al Khalij Commercial Bank	2007	2007
10	Barwa Bank	2010	2009
	Saudi Arabia		
1	Al Awwal Bank	2000	1926
2	The National Commercial Bank	2000	1953
3	Riyad Bank	2000	1957
4	Al Rajhi Bank	2002	1957
5	Bank AlJazira	2000	1975
6	Saudi Investment Bank	2000	1977
7	Banque Saudi Fransi	2000	1977
8	The Saudi British Bank	2000	1978
9	Arab National Bank	2000	1979
10	Samba Financial Group	2000	1980
11	Bank AlBilad	2005	2004
12	Alinma Bank	2009	2008
	United Arab Emirates		
1	National Bank of Dubai	2000-06	1963
2	Mashreq Bank	2000	1967
3	National Bank of Abu Dhabi	2000-16	1968
4	Commercial Bank of Dubai	2000	1969
5	Bank of Sharjah	2000	1973
6	Arab Bank for Investment & Foreign Trade	2000	1975
7	United Arab Bank	2000	1975
8	Invest Bank	2000	1975
9	Dubai Islamic Bank	2000	1975
10	National Bank of Ras Al Khaimah	2000	1976
11	First Gulf Bank	2000-16	1979
12	Emirates Bank International	2000-06	1979
13	National Bank of Fujairah	2000	1982
14	National Bank of Umm Al Qaiwain	2000	1982
15	Union National Bank	2000	1982
16	Abu Dhabi Commercial Bank	2000	1985
17	Commercial Bank International	2000	1991
18	Abu Dhabi Islamic Bank	2000	1997
19	Sharjah Islamic Bank	2002	2002
20	Emirates Islamic Bank	2004	2004
21	Emirates NBD Bank	2007	2007
22	Al Hilal Bank	2008	2008
23	Noor Islamic Bank	2011	2008
24	Ajman Bank	2010	2009
25	First Abu Dhabi Bank	2017	2017

Notes:

(1) For Bahrain:

- Ithmaar Bank was operating in Bahrain as an investment bank and converted into a retail bank in 2010.
- Shamil Bank merged into its parent company, Ithmaar Bank, in 2010.
- Al-Salam Bank acquired the Bahraini Saudi Bank in 2009 and acquired BMI Bank in 2014.

(2) For the UAE:

- National Bank of Dubai & Emirates International Bank merged in 2007 to create Emirates NBD Bank.
- National Bank of Abu Dhabi & First Gulf Bank merged in 2017 to create First Abu Dhabi Bank.

(3) For Saudi Arabia:

- Al Awwal Bank was formerly known as the Saudi Hollandi Bank and rebranded to Al Awwal Bank in 2016.

(4) The sample included is until the year 2017 unless stated otherwise.

Appendix B: Baseline model estimation results using MD-MLE

The table provides lower bound estimates of the baseline model for using a minimum distance maximum likelihood estimator. Standard errors are in parenthesis. The dependent variables ($\tilde{C}_1$ and $\tilde{C}_3$) are the logit transformation of CR_1 and CR_3, respectively. The independent variables are the two measures of market size (assets and population). Assets is the natural logarithm of total banking assets over set-up costs; Population is the natural logarithm of population over set-up costs; β_0 and β_1 are the estimated parameters and correspond to the intercept and slope, respectively; Ln a shape and Ln b scale are Weibull parameters; and C^{∞} is the limiting level of concentration reached at infinite market size.

	Assets		**Population**	
	$\tilde{C}_1$	$\tilde{C}_3$	$\tilde{C}_1$	$\tilde{C}_3$
β_0	-2.353***	-1.352***	-3.067***	-1.967***
	(0.0537)	(0.0732)	(0.0730)	(0.0587)
β_1	5.788***	8.688***	16.06***	18.83***
	(0.457)	(0.169)	(0.841)	(0.552)
Ln a shape	0.752***	0.928***	0.923***	0.746***
	(0.0840)	(0.166)	(0.0887)	(0.103)
Ln b scale	-0.179***	-0.477***	-0.259***	-0.445***
	(0.0565)	(0.108)	(0.0479)	(0.0623)
C_1^{∞} ; C_3^{∞}	9%	21%	4%	12%
95% Confidence Interval	[7.9; 9.5]	[18.2; 22.9]	[3.8; 5.1]	[11; 13.5]
Observations	108	108	108	108