Capital and profitability in banking: Evidence from US banks

MATTHEW OSBORNE\textsuperscript{1,2}, ANA-MARIA FUERTES\textsuperscript{2} and ALISTAIR MILNE\textsuperscript{3}

\textsuperscript{1}UK Financial Services Authority, \textsuperscript{2}Cass Business School, City University, London, \textsuperscript{3}Loughborough University School of Business and Economics

Abstract

This paper examines the effect of capital ratios on bank profitability over economic cycles using data from the US banking sector spanning several economic cycles from the late 1970s to the recent financial crisis of 2008-10. This relationship is likely to be time-varying and heterogeneous across banks, depending on banks’ actual capital ratios and how these relate to their optimal (i.e., profit-maximising) capital ratios, and we employ an empirical framework which allows substantial heterogeneity across banks and over time. While the average relationship across banks is negative for most banks in most years, it turns less negative or positive under distressed market conditions, namely the savings and loan crisis of the late 1980s and the recent global financial crisis of 2008-10. This is consistent with the hypothesis that in such conditions, increases in capital ratios are less costly for banks than in other periods. Since other factors may drive the long-run relationship between capital and profitability, we also examine the effect of short-run deviations from estimated long-run capital targets. Banks with a surplus of capital relative to target exhibit a strongly negative relationship between capital and profitability, both in stressed and non-stressed conditions, implying that reducing capital may be the optimal strategy for these banks. We conclude with policy implications, namely that counter-cyclical variations in capital requirements envisaged under Basel III will need to be large in order to achieve macroprudential aims of smoothing credit cycles.

\textit{Keywords:} Bank capital; Profitability; Bank regulation; Capital requirements.

\textit{JEL Classification:} G01, G18, G28, G32, G38.
1 Introduction

How does the level of capital affect profitability in banking? Higher capital is often supposed to be costly for banks, implying that higher capital reduces profitability, but according to the “trade-off” theory it may also reduce a bank’s risk and hence the premium demanded to compensate investors for the costs of bankruptcy. According to conventional corporate finance theories a bank in equilibrium will desire to hold a privately optimal level of capital that just trades off costs and benefits, implying a zero relationship at the margin. However, capital requirements imposed by regulators, if they are binding, force banks to hold capital in excess of their private optimal and hence force banks above their internal optimal capital ratio impose costs on banks (Miller, 1995; Buser et al, 1981). Furthermore, since banks’ optimal capital ratios are likely to vary over the cycle, typically rising when there are higher expected costs of distress, the relationship between capital and profitability is likely to be highly cyclical, becoming more positive during periods of distress as banks that increase their capital ratios provide reassurance to investors and improve their profitability. For example, the study of Berger (1995), to which our study is closely related, contrasts the positive relation between capital and profitability for the period 1983-89, a period of severe stress in the US banking system, with the negative relation found in 1990-92, when it is argued that banks may have exceeded their optimal capital ratios due to improved profitability and tighter regulatory capital ratios.

In this paper, we extend the results of Berger (1995) in order to show the effects of capital on profitability in the period 1993-2010, including the recent banking crisis of 2008-10. We then extent the model to consider further issues, namely whether there are asymmetries in the relationship between capital and profitability of US banks linked to deviations from banks’ own target capital ratios, and whether the relationship between capital and profitability varies over financial cycles. We use profitability as the measure of bank performance since it is a commonly used proxy, it is readily available for the banks in the sample and, unlike market value, it does not require assumptions about the market value of debt. We carry out our analysis using an exceptionally large dataset based on the US Federal Reserve Reports of Condition and Income (“Call Reports”), consisting of up to 15,000 banks over 30 years, in total more than 1.6 million bank-quarter observations.

The specific empirical contributions of this chapter are as follows:
First, it is the first study to chart in a systematic way the changing relationship between capital and profitability over such a long period of US banking history, including the recent banking crisis as well as the savings and loan crisis of the late 1980s. We also employ more recent econometric techniques for dynamic panel data.

Second, we embed a model of banks’ long-run target capital ratios \( k^* \) which recognises substantial heterogeneity over time and across different classes of banks in terms of size and portfolio risk, and use this to assess asymmetries in the relationship between capital and profitability depending on whether banks are above or below \( k^* \). This is more accurate than previous studies to have allowed non-linearity in the relationship which have divided banks into low and high capital in a statistical manner without taking into account banks’ internal capital targets (Osterberg and Thomson, 1996; Gropp and Heider, 2010).

The empirical analysis is in two parts. First, we extend the results of Berger (1995), a seminal paper which developed the idea that the relationship between capital and profitability may become positive during banking sector distress, namely the “savings and loan crisis” of the 1980s. This is ascribed to the “expected bankruptcy costs” hypothesis, an important component of the standard “trade-off” theory, which states that in stressed conditions higher capital reduces the required rate of return on risky debt by reducing the probability of default. We apply more recent econometric techniques to estimate the Berger (1995) model more robustly, and also extend the sample period up to 2010 to include the recent financial crisis. We confirm that the original findings of Berger (1995) are robust to new techniques and apply also, though to a lesser extent, in the recent financial crisis, since the negative relationship from capital to profitability found during the 1990s and early 2000s reduces significantly during this period. However, we note a number of shortcomings with the model which derive from the use of a reduced form specification to capture structural relationships. Hence, while the findings are attributed to the “expected bankruptcy costs” hypothesis, this is difficult to distinguish from other factors driving the long-run co-movement of capital and profitability. For example, a more profitable bank may choose to hold a lower capital buffer since it expects to be able to rely on internal funds to meet regulatory or market demands. Although this would usually drive a negative rather than a positive relationship, it is also possible that under stressed conditions, low profitability may lead a bank to increase leverage in order to rebuild profitability (Milne and Whalley, 2002).
Therefore, in the second part of the analysis we present initial results from an improved specification that attempts to address this problem and provide a more robust test of the “expected bankruptcy costs” hypothesis. We estimate long run target capital ratios for US banks, using a dynamic panel specification, and include these in an extended version of Berger (1995) in order to separate the effects of the long-run co-movement of capital and profits from short-run deviations. Using the estimated target capital ratios as a proxy for banks’ own internal optimal capital ratios, we assess the hypothesis that there are asymmetric effects of deviations from the optimal capital ratio. If a bank is below its target capital ratio, then we expect deviations (deficits) to be positively correlated with capital since banks can improve profitability by increasing capital, for example because investors are concerned about expected bankruptcy costs. If a bank is above its optimal capital ratio, then we expect deviations (surpluses) to be positively (negatively) correlated with capital since banks can improve profitability by reducing capital, consistent with the trade off theory. Since we remove the effects of the long-run co-movement of capital and profitability (via the inclusion of the target capital ratio itself in the regressions) our method is a more robust test of the trade-off theory for banks than the model based on Berger (1995).

The theoretical background is briefly as follows. Higher capital is often supposed to be costly for banks due to capital market imperfections and tax advantages of debt, but according to the popular “trade-off” view higher capital may also reduce risk and hence lower the premium demanded to compensate investors for the costs of bankruptcy. Therefore, there may be a positive or negative relationship between capital and firm value in the short run depending on whether a bank is above or below its optimal capital ratio. Indeed if banks are successful in attaining their optimal capital ratios there may in fact be no short-run relationship at all, since standard first order conditions imply that any change in capital has no impact on value. In the long run, regulatory capital requirements may exceed the bank’s optimal capital ratio and drive a negative relationship between capital and value, if they are binding. This implies that higher capital only reduces value if banks are above their optimal capital ratios, for example due to capital requirements or unexpected shocks.

One implication of this is that banks’ optimal capital ratios will rise during periods of banking sector distress (“bad times”), since in such conditions the expected costs of bankruptcy rise. Since capital market imperfections mean that banks cannot immediately adjust to the new optima, actual capital ratios tend to lag behind target capital ratios. Consequently, we expect that the average relationship between capital and value across banks
will be cyclical, since in a stressed environment banks tend to be below their optimal capital ratios, whereas during normal conditions (“good times”), banks may either meet their optimal capital ratios, in which case the relationship would be approximately zero, or overshoot, in which case banks can increase value by reducing the capital ratio (e.g., taking advantage of tax benefits of debt or implicit debt subsidies).

The long time period of our sample allows estimation of the relationship over several financial cycles. First, in the 1980s and early 1990s the US banking sector was in a state of upheaval due to the savings and loan crisis, which resulted in many bank failures and occurred simultaneously with a widespread recession, though empirical studies differ as to how far the recession was caused by banking problems (see reviews by Berger et al, 1995 and Sharpe, 1995). Second, throughout the 1990s and early 2000s capital ratios rose and profitability was consistently high. The higher capital ratios built up over this period have been attributed to the higher capital requirements of Basel I and the Federal Deposit Insurance Corporation Improvement Act (FDICIA) (Jackson et al 1999), increased market discipline due to the removal of explicit and implicit state guarantees, and passive earnings retention in a period of high profitability (Berger et al 2008; Flannery and Rangan, 2008). Finally in the late 2000s came the global financial crisis (GFC) which resulted in substantial losses and unprecedented levels of official support for the banking system.

In our study we use profitability as a measure of bank performance. Although this is an imperfect measure of bank performance, it is a commonly used proxy in the banking literature since it can be relatively easily calculated using accounting information and does not require assumptions to be made about the market value of a firm’s debt (see Mehran and Thakor, 2011 for a good discussion of this problem related to banks and an attempt to deal with it via the use of M&A data). Since the net income measure of profitability reflects interest paid on debt it is also sufficient to test our hypothesis of interest that expected bankruptcy costs drive a relationship between capital and value via the cost of funding. A possible future extension would be to consolidate individual banks and identify them with corresponding listed holding companies in order to test whether similar findings are observed also for market returns as for profitability.

A significant challenge in identifying the causal link from capital to profitability is that the direction of causality can plausibly run the other way, i.e. from profitability to capital. In the short run, high profitability may drive higher capital ratios since profits are a source of
capital. According to the pecking order theory, internal funds are the least information-intensive source of funds and hence a more profitable firm may retain earnings to fund known investment opportunities (Myers and Majluf, 1984). Hence, the bank’s capital is the sum of past profits less distributions. In the long run, a more profitable bank may desire a smaller capital buffer since it knows that it will be able to draw on internal funds to fund expected investment opportunities (Myers, 1984) or avoid regulatory censure (Milne and Whalley, 2002). High profitability may also affect the value of the tax deductability advantage offered by debt, since a firm that is not earning profits does not pay tax on payments to equity holders (Modigliani and Miller, 1958; Miller, 1977). These two factors indicate that more profitable banks will choose to hold lower capital ratios in the long run, i.e. a negative association. Milne and Whalley (2002) also point out that a bank that is distressed may exhibit a positive association since low profits elicit a gamble for resurrection involving higher portfolio risk and higher leverage. Alternatively, high profitability may lead managers to retain excess profits in order to fulfil their own personal projects or ambitions (Jensen, 1986). High profits may also increase the bank’s perceived charter value, providing an incentive to hold higher capital ratios (Marcus, 1984; Keeley, 1990). These factors predict positive causality from profits to capital in the long run.

Since they concern the long-run co-movement of capital and profitability, these effects cannot be completely removed by examining the relationship between capital and lagged profitability. Therefore, we exploit the large cross section dimension of our dataset in order more reliably to distinguish the effects of capital on profitability for banks which are above and below their long-run target capital ratios. We first estimate a long-run target capital ratio for each bank and quarter using a dynamic partial adjustment model (PAM) specification. The large cross-section dimension of our dataset is important since different banks will have very different target capital ratios in the long run; for example, Berger and Bouwman (2012) show that the extent to which higher capital ratios increase the performance of banks during stress episodes depends significantly on bank size. Consequently, we estimate target capital ratios separately for nine different groups defined in terms of their size and level of portfolio risk measured using a variety of indicators. We then again exploit the cross-section of our dataset by including deviations from the target capital ratio in a model of the return on assets (ROA), also controlling for the level of the target capital ratio itself.

Our findings are that in terms of the simpler model based on Berger (1995), the link from capital to profitability is strongly cyclical, since it is strongly positive during the 1980s crisis,
negative throughout the 1990s and early 2000s, and then becomes positive again in the recent financial crisis (though not to the same extent as during the 1980s crisis). The findings using the improved model based on target capitalization shed further light on these trends. Our results for the 1980s support the “expected bankruptcy costs” hypothesis proposed by Berger (1995) since we find a positive relationship for higher risk banks which are below their long-run target capital ratios. However, the same is not true for the recent financial crisis in 2007-2010 since there is no evidence for this period that low capital banks could improve profitability by increasing capital ratios. We suggest that this may reflect the role of official support in containing the effect of market discipline.

Our results also shed light on the driving forces behind the increases in capitalization during the 1990s, which has been the subject of several studies (Flannery and Rangan, 2008; Berger et al, 2008). From the basic model we learn that the link from capital to profitability was strongly negative for most banks, and our improved model shows that this was true for both low risk and high risk banks. There is only weak evidence of low capital banks being able to improve their ROA by increasing capital in this period. These findings are not consistent with an increased role for market discipline during this period, as proposed, for example, by Flannery and Rangan (2008), since this hypothesis would suggest that bank with low capital and high risk could increase profits by raising their capital ratios. Our results suggest that banks were on average well above their internal optimal capital ratios, and from another study we know that banks were also mostly well above the regulatory minimum capital requirement in this period (Berger et al, 2008). This may indicate that high profitability drove higher capital ratios during this period rather than the other way around.

We also use our results to draw policy implications. Our findings are relevant for a bank regulator seeking to calibrate capital requirements, and the short-run effects of capital on profitability are especially relevant for the design of so-called “macroprudential” tools such as counter-cyclical capital requirements.¹ Such tools are designed to dampen credit cycles by increasing the cost of new loans in a boom and stimulating lending during busts. Of course, since counter-cyclical capital requirements did not exist during the sample period, it is not possible to draw direct inferences about their effects from the historical data, and the indirect inferences we do draw may not apply to the same extent in a regime with much higher capital requirements that are binding on most banks (a regime that Basel III may deliver). However,

¹ For fuller reviews of macroprudential policy tools, see Milne (2009) and Bank of England (2011).
our results on how the relationship co-moves with financial cycles provides valuable information about the likely effects of counter-cyclical capital requirements, since to be effective, banks must have incentives to change their capital ratios and lending policies in response to cyclical adjustments in capital requirements. Our results indicate that banks’ own optimal capital ratios are cyclical, and since banks can increase their profitability by increasing their capital ratios during periods of financial distress such as the 1980s or late 2000s, it is not likely that they would respond to reductions in capital requirements in such conditions. During the late 2000s crisis the effects of market discipline were not as apparent as during the 1980s crisis. During the 1990s, capital ratios were rising anyway so it seems unlikely that increasing capital requirements in line with macroprudential aims would have affected banks’ capital choices. There were substantial downward pressures on banks’ capital ratios in this period, but only for banks with high capital relative to their long-run target capital ratios, which are the banks least likely to be affected by capital requirements.

This paper is structured as follows. In section 2 we review relevant literature on the theoretical relationship between capital and profitability, and also review of trends in capital and profitability in the US banking sector between 1976 and 2010. In section 3 we set out the first part of our analysis, an extension of Berger (1995). In section 4 we set out the second part of our analysis in which we present initial results of an improved model designed to more robustly test our hypotheses. Section 5 concludes.

2 Literature review

2.1 Theory and evidence on the relation between capital and profitability in banks

According to the second proposition of Modigliani and Miller (1958), investors’ required return on market equity is a negative linear function of the ratio of equity to debt, since higher leverage raises the return demanded by shareholders. Most academic studies have argued that deviations from the M-M theorems are particularly relevant for banks, and therefore banks have an optimal capital ratio which maximises their value (Buser, Chen and Kane, 1981; Berger et al, 1995). At the core of this literature is the theory that the tax advantages of debt and the advantages of government guarantees of debt “trade off” against the expected bankruptcy costs associated with low equity.
The effects of market discipline constrain banks to limit their leverage given that investors are sensitive to the default risk of the bank, implying that banks with a high probability of distress may be punished with a relatively high cost of uninsured funding (Nier and Baumann, 2006; Flannery and Rangan, 2008; Berger, 1995; Flannery and Sorescu, 1996; Covitz et al, 2004; Jagtiani et al, 2002; Morgan and Stiroh, 2001, Flannery, 1998; Sironi, 2003; and Gropp et al, 2006). Flannery and Rangan (2008) point out that the effect of market discipline is muted by government guarantees, and therefore may be primarily observed in periods where government guarantees are perceived to have been withdrawn, such as in the early 1990s in the US when the FDICIA limited protection of uninsured bank creditors from default losses and the Omnibus Budget Reconciliation Act of 1993 subordinated non-deposit claims to a failed bank’s deposits. A result which is at odds with this hypothesis is that of Covitz et al (2004) which contends that the lack of observed significant market discipline effects before this are the result of sample selection bias, since issuance is partly a function of the potential spread, and after correcting for this bias they find a significant effect of default risk on subordinated loan spreads throughout the period 1985-2002. It therefore seems likely that higher portfolio risk increases the optimal capital ratio of a bank, especially when the bank is subject to a greater degree of market discipline. An upwards shift in portfolio risk results in a more negative relationship between capital and profitability in the short-run, until the bank adjusts to its new optimal capital ratio.

The positive effect on capital ratios of market discipline from liability holders may be reinforced by the incentives of the bank’s owners and/or managers who also stand to lose from the failure of the bank given that they will then lose future rents from the bank. The charter or franchise value of a bank is the net present value of future rents which accrue to the owners or managers, such as larger interest margins arising from market power or established relationships, and provide incentives for banks to limit banks’ risk taking at the expense of liability holders (Marcus, 1984; Keeley, 1990; Demsetz et al, 1996; Hellmann et al, 2000; Bhattacharya, 1982; Rochet, 1992). Hence, higher profitability creates incentives for banks to limit risk-taking, including holding higher capital ratios as well as reducing portfolio risk.

One implication of this that has been explored by a number of recent studies is that higher capital may help overcome agency problems arising from information asymmetries in the bank-investor and bank-borrower relationships (Holmstrom and Tirole, 1997; Mehran and Thakor, 2011; Allen, Carletti and Marquez, 2011). A bank with a higher capital ratio has more chance of surviving in the future and therefore has a greater incentive to monitor
borrowers, and investors take this into account when valuing claims on the bank. Allen, Carletti and Marquez (2011) argue that this may result in large voluntary capital buffers in competitive markets, since then higher capital is a more effective guarantee of the bank’s monitoring incentives and therefore allows the bank to offer more surplus to borrowers. The effect is to increase banks’ optimal capital ratios.

In the opposite direction, the effect of explicit or implicit government guarantees of banks’ liabilities tend to reduce optimal bank capital. In particular, deposit insurance schemes weaken the disciplining effect of depositors’ required rate of return (Merton, 1978). Perceptions amongst investors that the government will guarantee liabilities upon default achieves a similar effect for uninsured debt (O’Hara and Shaw, 1990; Nier and Baumann, 2006; Flannery and Rangan, 2008). The effect of a greater perceived likelihood of a safety net being in place is to reduce optimal bank capital ratios.

Another plausible determinant of the optimal capital ratio is the relative information costs associated with different sources of funding. According to the pecking order theory of financing, raising equity or debt is costly for the bank because it conveys via a signal to the market the owners’ or managers’ view of the prospects of the bank (Myers and Majluf, 1984). In the pecking order theory a firm’s capital ratio is determined by the availability of internally-generated funds from the past and the firm’s investment opportunities. This may generate an optimal capital ratio since the firm may also be concerned with maintaining financial slack in order to be able to take advantage of future investment opportunities (Myers, 1984). A highly profitable bank may therefore maintain a lower capital ratio since it expects to be able to fund future investment out of earnings.

Banks’ choices of capital ratio may also depend on their business plans. A bank with an aggressive business strategy aimed at gaining market share may leverage up rapidly and hence has a lower capital ratio. A bank that plans to acquire another bank may have an incentive to maintain a higher capital ratio, for example so that it can satisfy regulators that the resulting entity will be adequately capitalised (Berger et al, 2008). A bank that is increasing its market share may hold a lower capital ratio consistent with a higher risk strategy or simply because loan growth runs ahead of the ability to raise or retain capital (Goddard et al, 2004).

A further important determinant of banks’ capital ratios is regulation, in the form of explicit minimum capital requirements and other supervisory pressures related to capital or
profitability. Such regulation is equivalent to an additional cost that is decreasing in the level of capital that the bank holds. For banks that fall below the required minimum level of capital it could include suspension of permissions, replacement of management or the imposition of tough plans to restore capital adequacy, implying a positive probability that the bank is closed so that its owners lose valuable franchise value (Merton 1978; Bhattacharya et al 2002). Hence, several studies have analysed a trade-off between the benefits of holding lower capital ratios and the expected costs of regulatory intervention, predicting that banks optimally hold a buffer of capital over regulatory capital requirements (Estrella, 2004; Milne and Whalley, 2002; Peura and Keppo, 2006; Barrios and Blanco, 2003; Repullo and Suarez, 2008; Heid, 2007). Following the same logic, the optimal capital buffer is likely to increase in the degree of portfolio risk, though identifying this effect is made difficult by the fact that risk averse banks may have higher capital and lower portfolio risk, and in the short run banks with low capital buffers may “gamble for resurrection” with higher risk (Lindquist 2003; Jokipii and Milne, 2011; Peura and Keppo, 2006). In addition, the choice of capital buffer may also be a function of profitability, since a highly profitable bank expects to be able to draw on internal funds to protect against falling below the regulatory minimum (Milne and Whalley, 2002), a similar argument to the pecking order theory discussed above.

Hence, capital requirements are likely to be an important influence on the bank’s capital choice, at least if they are binding. If capital requirements are binding they may force a bank to hold capital above the value-maximising level (as determined by non-regulatory factors described above), implying a negative long-run and short-run relationship between bank capital and profitability. If they are not binding on the bank, the relationship may be positive, flat or negative as it would be in the absence of capital requirements. Whether capital requirements are binding depends on the level at which they are set and the capital ratios that the bank would choose in the absence of capital requirements, suggesting variation across countries, banks and time periods. If banks’ own optimal capital ratios are cyclical, the impact of flat capital requirements will be cyclical as well (Blum and Hellwig, 1995; Heid, 2007). Perhaps reflecting this ambiguity, the empirical literature on whether capital requirements affect banks’ capital ratio decisions is mixed. Some studies find a positive link (Francis and Osborne, 2010; Osterberg and Thomson, 1996), while a number of other studies cite large buffers held by banks as evidence that capital requirements are not binding (Gropp and Heider, 2010; Berger et al, 2008; Flannery and Rangan, 2008).
However, large buffers do not by themselves constitute evidence that capital requirements are not binding. Investors may internalise the expected costs of regulatory intervention, which are always positive as long as there is a positive probability of falling below the capital requirements, but which are likely to be much greater the closer the bank is to the regulatory minimum (Milne and Whalley, 2002; Jokipii and Milne, 2011; Gropp and Heider, 2010). Large buffers may therefore reflect high portfolio risk, expectations about future business opportunities, or temporary deviations of the buffer from the previous level due to unexpectedly high profitability or low asset growth. This calls into question the approach adopted by some studies to assume that banks are either bound by capital requirements, in the sense that their own optimal capital ratios are below the regulatory required level, or not, meaning that their optimal capital ratios exceed the regulatory minimum (Wall and Peterson, 1988; Barrios and Blanco, 2003).

In addition, capital requirements may affect the market’s view of the level of capital for a bank. Assuming that supervisory assessments and related pressures on capital are to some extent publicly available and represent more information than is available to the market (DeYoung et al, 2001; Berger et al, 2000), capital requirements may also alter investors’ perception of the capital ratio that is necessary to assure the solvency of the bank, even for a bank that has a sizeable buffer of capital. Hence, the effects of market discipline are closely interrelated with capital requirements (e.g., see Jackson et al, 1999; Osterberg and Thomson, 1996). Since investors penalise a bank with insufficient capital with a higher required rate of return, the effect of capital requirements is that banks with capital ratios close to the regulatory minimum are likely to have a more positive relationship between capital and profitability than those that are less constrained, other things equal.

The discussion above offers a number of competing drivers of banks’ capital ratios which suggest that the relationship could be positive or negative depending on banks’ circumstances. Most empirical studies that have examined the relationship between capital and profitability have found a positive relationship, across a variety of different markets and time periods (Angbazo, 1997; Demirguc-Kunt and Huizinga, 1999; Vennet, 2002; Nier and Baumann, 2006; Flannery and Rangan, 2008). This could be attributed to the pecking order theory since high earnings in the past drive higher capital in the present, although, as noted above, both the buffer capital and pecking order perspectives predict that in the long run higher profitability leads to a lower desired capital ratio. Morgan and Stiroh (2001) look
directly at the relationship between capital ratio and spreads paid on debt and find a relationship that is positive but weak.

However, the studies cited above have a limitation since they impose the restriction that the relation between capital and profitability is linear and homogeneous across banks and time periods. The positive relationship found is therefore an average relationship across banks and time periods. We are aware of three studies that have tried to relax this constraint. Berger (1995) estimates a linear relationship between ROE and lagged capital for US commercial banks but allow the relationship (and in fact the whole model) to vary between the periods 1983-89, in which banks in general are argued to have been below their optimal capital ratios, and 1990-92 in which they are believed to have been above their optimal capital ratios. The coefficient is positive in the first period and negative in the second, consistent with the theory above that optimal capital ratios rise during periods of distress and banks depart from their optimal capital ratios in the short run. However, this method does not allow the relationship to vary across banks, which from a theoretical point of view it is likely to do depending on whether banks are above or below their optimal capital ratio and how close they are to it.

Osterberg and Thomson (1996) explore the determinants of leverage in US banking holding companies (BHCs) in the period 1987-8. Their specification includes a linear earnings term and an interaction term consisting of earnings multiplied by the buffer over capital requirements. They find that the capital ratio and earnings are positively correlated, and this relationship is more positive for banks which are close to the regulatory minimum. Gropp and Heider (2010) study the determinants of leverage for large US and European banks over the period 1991-2004. They include return on assets and also the return on assets multiplied by a dummy variable which is equal to 1 if the bank is close to its regulatory requirement. The ROA and capital ratio are negatively related, but when capital is close to the regulatory minimum, the overall relationship is insignificantly different from zero.

The non-linear specifications used in Gropp and Heider (2010) and Osterberg and Thomson (1996) suffer from a drawback since the relationship can only vary for banks with low capital. The factors driving the relationship between capital and profitability identified above point to a relationship that could differ depending on whether a bank is above or below the optimal capital ratio and therefore the relationship could vary between high capital and medium capital banks as well as for low capital banks. A related issue is that the studies attribute the non-linearity close to the regulatory minimum to the effects of regulation. While
it is true that regulation may drive a particularly strong positive relationship close to the regulatory minimum, market discipline could drive similar effects and it is questionable whether those studies have truly identified an incremental regulatory effect distinct from the trade-off factors which are relevant for any firm.

In section 3, we set out a specification for the relationship between capital and profitability which allows for heterogeneity across time and interaction with the level of capital across banks. We apply this specification to an updated dataset including commercial banks through the recent financial crisis.

2.2 Trends in capital and profitability of US banking sector, 1976-2010

As documented by Berger et al (1995), for most of the 19th century the US banking sector had an equity-to-assets ratio exceeding 30%, but by the 1950s the ratio had fallen to 7-8%, and fell further to around 5% by the 1980s. During the late 1980s the sector went through a period of profound stress, with large loan losses and annual bank failures numbering in the hundreds. In the same period, banks’ charter values were also eroded by a loss of market power due to deregulation such as the repeal of deposit rate caps and reduction of inter-state banking restrictions (Berger, 1995; Marcus, 1984). In Figure 2.1 below, we show trends in the capital-asset ratio of US commercial banks, showing separately the 10th, 50th and 90th percentiles and the mean, and in Figure 2.2 we show the same statistics for large banks only, defined as those with assets in the top decile each year. The k stayed fairly level in the 1980s before rising sharply through the 1990s, dipping only briefly in 2000-01, and then maintaining its rise until the late 2000s GFC when it dips noticeably (Figure 1, panel A). The same trends are evident for large banks alone (Figure 2, panel A). The stress of the mid-1980s is clearly evident in the ROE for all banks (Figure 1, panel B), which declines across the whole distribution, reaching the lowest point in 1986 before recovering in 1987-92. For large banks, the decline in ROE (Figure 2, panel B) in the mid-1980s is less marked and does not appear at all for banks at the 50th or 90th percentiles, suggesting that this crisis largely affected small and medium sized banks. It is possible that the 1980s shocks could reflect local factors such as property price falls, which would mainly affect local banks with regionals.

---

2 We do not attempt a full review of trends in the US banking industry here, preferring to identify only those facts that are most salient to our analysis. For a more complete review see DeYoung (2010).
specific loan exposures. The downwards effect of the late 2000s GFC on ROE is evident across all banks at all points of the distribution.

Figure 2.1: Descriptive statistics on US commercial banks' capital-asset ratio and return on equity, 1976-2010, all banks

Panel A: Capital-asset ratio

Panel B: Return on equity

Source: US Federal Reserve Reports of Condition and Income (Call Reports). Capital-asset ratio defined as the ratio of equity to assets. Return on equity defined as net income over equity. Extreme observations (defined as those with k or ROE over 3 standard deviations from the mean) are excluded.
Figure 2.2: Descriptive statistics on US commercial banks' capital-asset ratio and return on equity, 1976-2010, large banks

Panel A: Capital-asset ratio

Panel B: Return on equity

Source: US Federal Reserve Reports of Condition and Income (Call Reports). Large banks are defined as those in the top third in terms of total assets in each quarter (groups 7-9 in the analysis below). Capital-asset ratio defined as the ratio of equity to assets. Return on equity defined as net income over equity. Extreme observations (defined as those with k or ROE over 3 standard deviations from the mean) are excluded.
Partly as a result of the experience of the 1980s, the period from 1989-93 saw a considerable tightening of regulatory standards. These measures had been foreshadowed by the introduction of a leverage ratio requirement in 1984, which required ‘primary’ capital (equity plus loan loss reserves) to be more than 5.5% of assets and the total of ‘primary’ capital and ‘secondary’ capital (primarily qualifying subordinated debentures) to be over 6% of assets, which seems to have been effective in raising banks’ capital ratios (Wall and Peterson, 1988). The forthcoming risk based capital requirements had also been flagged in the publication of a joint US/UK proposal on risk-based capital (RBC) requirements in 1986 (Osterberg and Thomson, 1996). The increase in capital-asset ratios of large banks in 1985-89 evident in Figure 2.1 may reflect these measures, though there is no similar trend for small banks.

Risk-based capital requirements were formalised in the international Basel Accord of 1988, and then introduced in the US between 1990 and 1992, and a higher leverage ratio requirement was also introduced in 1990 (Berger et al, 1995). The leverage ratio may have been the binding constraint for many banks whose portfolios contained mainly low risk-weighted assets such as mortgages and government securities (Berger and Udell, 1994). At the same time, the FDICIA introduce prompt corrective action which increased supervisory sanctions for breaching regulatory minimum capital ratios, and also introduced risk-based deposit insurance premia. The combination of these regulatory measures incentivised banks to raise their capital ratios (Berger et al, 1995) and we can clearly see the effects of this in Figure 2.1 and Figure 2.2 which show a substantial increase in capital ratios in the early 1990s for large and small banks, and across the whole distribution.

The new regulatory standards were accompanied by a sharp increase in the capitalization of US banks and a credit crisis and a recession across much of the US, suggesting that adjustment to higher standards induced deleveraging and a contraction in banks’ supply of credit to the real economy, exacerbated by the relatively high risk weights given to corporate lending compared to government debt (Berger et al, 1995). However, the large literature on this subject provides rather mixed conclusions, since there is not a consistent association between closeness to the regulatory minima and changes in lending patterns at individual banks (e.g. see reviews in Berger et al, 1995; Sharpe, 1995; Jackson et al, 1999). This suggests that while regulatory changes may have played a role, much of the decline in credit can be attributed to a recession-induced contraction in demand.
A notable feature of Figure 2.1 and Figure 2.2 is that capital ratios kept increasing throughout the 1990s and into the early 2000s, for both large and small banks. While this does not rule out that the regulatory standards of the early 1990s were partly responsible for the increase, it does suggest that other factors were at play, since by the early 2000s capital ratios reached levels far in excess of the regulatory minima (Berger et al, 2008; Flannery and Rangan, 2008). One explanation for this trend is enhanced market discipline. In the 1980s, bank creditors enjoyed substantial guarantees effectively making large banks “too big to fail”, but a number of regulatory changes were made in the 1980s and 1990s which had the effect of reducing uninsured liability holders’ perceptions of the likelihood that they would be rescued upon default of a bank, including purchase and assumption transactions, prompt corrective action, least cost resolution and depositor priority (Covitz et al, 2004; Flannery and Rangan, 2008). Combined with much higher asset volatility over this period (driven in part by the 1997 Asian crisis and 1998 Russian default/LTCM problems), it has been argued that increased market discipline raised banks’ optimal capital ratios well in excess of capital requirements (Flannery and Rangan, 2008). Berger et al (2008) also note that that banks’ profitability rose in the early 1990s and remained at high levels into the mid-2000s, which is also evident from our Figure 2.1 and Figure 2.2, and this may have contributed to higher capital for reasons explained in section 2.1 above.

A number of other trends were evident in the 1990s onwards. The risk-weighted capital ratios stipulated by the Basel Accord fell, even as the unweighted ratio continued to rise (Berger et al, 2008). Part of the explanation for this phenomenon is the sector’s adaption to the imperfect risk-sensitivity in the rules. For example, a favourable treatment of securitisation under Basel I may have contributed to the growth of banks’ involvement and the pervasive “originate-to-distribute” business model in this period, although certain early studies failed to find a direct link between capital adequacy and the decision to securitize assets (Carlstrom and Samolyk, 1995; Jagtiani et al, 1995). Thirdly, this period saw substantial consolidation in the banking sector, due in part to deregulation noted above, particularly the removal on restrictions on interstate banking and on combining different banking activities such as investment banking with traditional banking activity (Berger et al, 1999).

Last but certainly not least, our sample period ends with the global financial crisis of 2007-09. The crisis was triggered by losses on subprime loans in the US which banks had tranched and sold on to a diverse range of investors including other banks, causing a contagious loss of
confidence in the balance sheets of banks and a collapse in interbank lending (Brunnermeier, 2009). The underlying causes of the crisis probably included global macroeconomic imbalances and low interest rates (see Merrouche and Nier, 2011 for a review of this emerging literature) and continue to be debated (see Lo, 2012). There are several implications for our consideration of capital and profitability. First, banks suffered substantial losses due to bad credit and were also forced to pay much higher rates of return on short-term and unsecured debt, contributing to much lower profitability over this period. Second, the capital ratios of banks have been a key determinant of profitability and survival during the crisis (Berger and Bouwman, 2012). As a result, the crisis has been associated with regulatory demands for much higher capital ratios including in the Dodd-Frank Act and international Basel III standards to which the US is a signatory. Significantly, the Basel III package includes calls for a leverage ratio based on the ratio of capital to assets, which, as noted above, has been a feature of US capital regulation since 1990.

The financial crisis of 2008-2010 has also seen unprecedented levels of official support for the banking sector, including emergency liquidity support such as the Troubled Asset Relief Program (TARP) which was launched in late 2008 to purchase assets from banks and later directly purchased the equity of banks, and other programmes operated by the Federal Reserve. Monetary policy was very loose by historical standards and fiscal policy was also used to stimulate the economy. Perhaps as a result of these dramatic actions, the actual rate of bank failure was much lower than the earlier savings and loan crisis in the late 1980s. Figure 2.3 shows the number of annual bank failures or assistance transactions recorded in each year from 1976 to 2010 by the Federal Deposit Insurance Corporation (FDIC) (dashed line, left axis), and the assets of failed or assisted banks as a percentage of total banking assets (full line, right axis). The number of failed banks reached 400-500 per year in the late 1980s, but in the recent GFC it peaked at only 150 per year. This may partly reflect the fact that there were many fewer banks in the industry at the time of the GFC. It also reflects the fact that failures were much more concentrated in relatively few large banks, which can be seen from the fact that assets of failed banks peaked at about 16% of total banking assets in 2009\(^3\) compared with only 5% in 1988.

\(^3\) The biggest bank that failed or received assistance from the FDIC in 2009 was Bank of America whose assets are 11% of the total. The next biggest in that year are FIA Card Services, part of Bank of America (1.2%) and Countrywide (0.9%).
Figure 2.3: FDIC bank failures 1976-2010

Notes: FDIC failed banks database was used to identify banks which failed or were assisted in each year. The assets of failed banks from this database are expressed as a percentage of total banking assets calculated from Federal Reserve Reports of Conditions and Income (call reports).

3 Extending the results of Berger (1995)

The first stage of our analysis is to extend the results of Berger (1995). As noted above, this paper’s important contribution was to identify that the effect of capital on the profitability of banks is likely to be cyclical. During the period 1983-89, which was a period of significant stress in the US banking sector (the “savings and loan crisis”), lagged values of the $k$ are found to be positively correlated with ROE, confirming a Granger causal relationship. It is argued that, during this period, banks’ optimal capital ratios rose substantially, and “banks that raised their capital toward their new, higher equilibria had better earnings than other banks through lower interest rates paid on uninsured debt. This is because higher capital reduces the probability that uninsured debt holders will have to bear the administrative, legal, and asset devaluation costs of bank failure, and therefore lowers the required premium on uninsured debt for banks that increase their capital ratios.” The same model is re-run during the early period of 1990-92, by which time banks’ capital ratios had increased substantially as shown in Figure 2.1 above. A negative relationship is found for this period. Berger
concludes, “The negative causality during this later period suggests that banks may have overshot their optimal capital ratios, as reduced risk lowered optimal capital ratios and regulatory changes and higher earnings raised actual capital ratios”.

Our first goal is to find out whether the original results of Berger (1995) stand up to more modern econometric examination, and whether the estimated positive relationship in stressed conditions also applies during the more recent period of banking distress. More specifically, we extend the results of Berger (1995) in several ways.

- We extend the sample period to the recent financial crisis, estimating the model on five-year rolling windows from 1977-1981 to 2006-2010. Using rolling windows is intended to capture smoothly changes in the relationship over time, in order better to uncover cyclical variations.
- We address the estimation problem of lagged dependent variable bias which is likely to reflect the earlier estimates, using more recently developed econometric techniques based on the general method of moments.
- We introduce controls for the banks’ portfolio risk, including the standard deviation of ROE which we argue is a more general measure of risk than the measures used by Berger (1995) and may better capture banks’ off-balance sheet, investment banking and trading activities.

3.1 Data and specification

The basic model used in Berger (1995) is summarised by Equation (1). \( ROE \) is regressed on three lags of itself and of \( k \), controlling for other potential determinants of profitability in the vector \( X_{i,t-1} \) and bank and time effects. We estimate Equation (1) on a set of 30 5-year rolling windows which we denote \( \tau \), starting in 1977-1981 and ending in 2006-2010. Hence, each set of coefficient estimates are specific to a window \( \tau \).

\[
ROE_{it} = \alpha_t + c_{it} + \theta_{tt} + \sum_{j=1}^{j} \beta_{j}^{1} k_{i,t-j} + \sum_{j=1}^{j} \beta_{j}^{2} ROE_{i,t-j} + \beta_{j}^{3} X_{i,t-1} + \gamma_{t} SIZE_{it} + \varepsilon_{it}
\]  

(1)
This model deals with causality from ROE to $k$ by including $J$ lags of the two dependent variables, which is equivalent to assuming that $k$ is predetermined, i.e. that while $ROE_{it}$ may affect $k_{it}$, $ROE_{it}$ does not affect $k_{i,t-1}$ since profitability would not affect past values of $k$. We calculate the effect of lags of $k$ and $ROE$ on $ROE$ by summing the coefficient over the $J$ lags, i.e. $\beta^1 = \sum_{j=1}^J \beta_j^1$ and $\beta^2 = \sum_{j=1}^J \beta_j^2$.

The vector of control variables $X_{it}$ include variables which previous studies have found helpful in explaining both capital and profitability. The coefficients on $CAR$ may then be interpreted as the effect of deviations from the level of capital that would be expected given the bank- and time-specific control variables, including bank and time effects (see Berger, 1995, p.440). Hence, the control variables are chosen to include both variables helpful in explaining leverage and those found helpful in explaining ROE (there is likely to be substantial overlap). We include bank- and time effects and also nine dummy variables representing deciles of the distribution of total assets in each year (SIZE). We include controls for the level of portfolio risk. These are expected to be positively correlated with ROE since higher risk is linked to higher returns for shareholders. Portfolio is also expected to be positively correlated with $k$ since investors trade off capital and risk in order to target a given probability of default (Jokipii and Milne, 2011). We include a set of different measures of portfolio risk identified by previous studies (e.g. Osterberg and Thomson, 1996; Jokipii and Milne, 2011; Berger, 1995), each capturing a separate aspect of risk:

**Ratio of commercial loans to total assets.** Commercial lending is generally acknowledged as the highest risk type of loans and consequently was given a 100% risk weight in the first Basel Accord. This is available for the whole sample period.

**Ratio of net charge-offs to total assets.** Charge-offs are a good ex-post measure of risk since they indicate the level of losses the bank has experienced. We construct a measure of annualised charge-offs and then, since charge-offs tend to be very volatile, we calculate a moving average over three quarters. This is available for the whole sample period.

**Ratio of non-performing loans to total assets.** Indicates the level of likely future losses on loans. This is available from 1984 onwards.

**Ratio of risk-weighted assets to total assets.** Introduced by the Basel Accord in 1990, this variable serves as an ex-ante measure of a bank’s risk. It has the advantage
that it captures the risk of trading book activities and off-balance sheet commitments. This series starts in 1996 (before then it does not include trading book assets).

**Standard deviation of return on equity.** From an investor’s point of view a key measure of the risk of a bank is the volatility in profitability (Flannery and Rangan, 2008). We calculate the standard deviation of the quarterly book return on equity over the last three years. This has the advantages that it does not rely on any ex ante weighting of different asset types (as do RW and C&I) and it captures volatility in profits as well as level of losses (which are captured by CHARGE, NPL). It also captures off-balance sheet and trading activities. This variable is available for the whole sample period.

We also include proxies for the competitive position of a bank’s markets, which is likely to be a key determinant of profitability according to the structure-conduct-performance framework (e.g., see Goddard et al, 2004). As described above, the US banking sector has become much more concentrated over the period of our analysis, though the removal of restrictions on interstate banking and technological advances may also have made the market more competitive at a local level (Berger et al, 1999). Therefore, we also include measures of market structure which are based on those used by Berger (1995) and Berger et al (1999). In order to capture local market concentration, we first identify local banking markets as Metropolitan Statistical Areas (MSAs) and non-MSA counties. The FDIC’s annual Summary of Deposits data gives details of each bank’s deposits in each local market. We calculate the bank’s share of local market deposits (SHARE), the Herfindahl index for each market (HERF), and the growth of deposits in each market-year combination (GROWTH). For each bank, we calculate the weighted average of each of these variables over the local markets in which it operates, with weights given by the share of its deposits in each market. We also calculate the share of a bank’s deposits in urban markets (i.e. MSAs).

Finally we also include the ratio of operating costs to total assets averaged over the last two years (AC), as a proxy for operating efficiency. This variable is only available from 1984 onwards.

---

4 Note that Berger (1995) includes RWA going back all the way to 1982, based on proxy data for balance sheet components. We were unable to obtain this data from public sources.

5 Our preference would be to use the return on market equity rather than book equity, but these are not available for our dataset of commercial banks.
In order to be consistent with Berger (1995), we first estimate Equation (1) using fixed effects. However, we note that this estimation strategy suffers from lagged dependent variable (LDV) bias since the fixed effect is correlated with the LDV, $\sum_{j=1}^{3} \text{ROE}_{t-j}$. Therefore, we also adopt the system GMM method developed by Blundell and Bond (1998) to estimate Equation (1). In this approach, the differenced LDV is instrumented with lagged levels, and the lagged LDV in levels is instrumented using lagged first differences. This estimator is available in one-step and two-step variants. While the two-step is asymptotically more efficient, the standard errors are downward-biased and must be corrected using the adjustment developed by Windmeijer (2005).\(^6\) We estimate both one-step and two-step versions of Equation (1). The number of lags of $k$ and ROE, $J$, is set at 3 consistent with Berger (1995).

The data are taken from the quarterly US Federal Reserve Reports of Condition and Income (Call Reports). We drop failed banks in the year in which they fail (using an FDIC database of bank failures) and banks under “special analysis” such as bankers’ banks, credit card banks, depository trust companies, and bridge entities. When $k$ falls below 1%, equity is replaced with 1% of assets for calculation of other variables with equity in the denominator. This avoids extreme and implausible values when equity is very low. Equity and assets are calculated as the average of the current value and three (quarterly) lags.\(^7\) Finally, we drop extreme observations of $k$ and ROE, defined as those which are greater than three standard deviations from the mean in each year. In total, our sample has just under 15,000 banks at the beginning of the sample period in 1977, falling to under 11,000 by 1995 and around 7000 by 2010. Our estimations are based on annual data using the fourth quarter in each year, a total of 30 time periods. The means of the control variables are shown in Table 3.1, separately for selected 5-year windows.

---

\(^6\) All of this is accomplished using the very useful *xtabond2* program developed in Stata by David Roodman (2004).

\(^7\) These steps to clean the data are based on those noted in Berger (1995) and Berger et al (2004).
Table 3.1: Descriptive statistics for control variables, 1977-2010

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BANKSHARE</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.15</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>HERF</td>
<td>0.26</td>
<td>0.25</td>
<td>0.25</td>
<td>0.23</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>GROWTH</td>
<td>0.12</td>
<td>0.11</td>
<td>0.04</td>
<td>0.07</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>MSA</td>
<td>0.41</td>
<td>0.44</td>
<td>0.45</td>
<td>0.46</td>
<td>0.51</td>
<td>0.53</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>CHARGE</td>
<td>0.001</td>
<td>0.003</td>
<td>0.004</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>C&amp;I</td>
<td>0.11</td>
<td>0.13</td>
<td>0.12</td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>SDROE</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>NPL</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td></td>
<td>0.64</td>
<td>0.67</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIZE</td>
<td>0.50</td>
<td>0.51</td>
<td>0.51</td>
<td>0.50</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Summary statistics for the standard deviation of ROE (SDROE), the ratio of commercial and industrial loans to total assets (C&I), the ratio of charge offs to total assets (CHARGE), the ratio of non-performing loans to total assets (NPL), the ratio of risk weighted assets to total assets (RW), the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index of concentration in the bank’s local markets (HERF), the share of a bank’s deposits in urban markets (MSA), the growth of local market deposits (GROWTH), the ratio of operating costs to total assets (AC) and size index (SIZE).

3.2 Results

We show the results of our estimation of Equation (1) in Figure 3.1 below. Panel A shows the fixed effects results, which are the most consistent with the method used by Berger (1995). Panels B and C show the results using one-step and two-step GMM respectively. Consistent with that paper’s findings, the effect of $k$ on ROE turns strongly positive in the late 1980s, before turning negative in the early 1990s. These findings are consistent across all three estimators, showing that the original Berger (1995) findings are robust to more modern econometrics. After the Berger (1995) sample period, we show that the negative relationship from $k$ to ROE persists throughout the 1990s. The FE estimates show that the negative relationship continued through the late 1990s and into the early 2000s, whereas the more robust GMM estimates show that the negative effect was close to zero by the late 1990s. In the late 2000s, with the onset of the GFC, all three estimators show the relationship
becoming markedly less negative, though only the GMM estimators find that it is significantly positive. These findings provide support for the idea that banks that increased capital ratios before and during the GFC had higher ROE, other things equal (although we note various concerns with this specification in section 3.3).
Figure 3.1: Estimated effect of $k$ on ROE 1977-2010

Panel A: Fixed effects

Panel B: One-step system GMM

Panel C: Two-step system GMM

Notes: Shows the results of estimating Equation (1) on annual data over 30 5-year rolling windows from 1977-1981 to 2006-2010. $k$ is the capital-asset ratio, ROE is the return on equity. We include bank and time effects, and also control for the standard deviation of ROE (SDROE), the ratio of commercial and industrial loans to total assets (C&I), the ratio of charge offs to total assets (CHARGE), the ratio of non-performing loans to total assets (NPL), the ratio of risk weighted assets to total assets (RW), the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index of concentration in the bank’s local markets (HERF), the share of a bank’s deposits in urban markets (MSA), the growth of local market deposits (GROWTH), the ratio of operating costs to total assets (AC) and 9 size dummies. Shaded bands show two standard errors (i.e. approximately the 95% confidence interval).
3.3 Problems with the model based on Berger (1995)

While Berger (1995) is an important contribution in the sense of identifying the idea that capital may have different effects on bank performance at different stages of the financial cycle, there are a number of problems with the above model, which we discuss below. The last two of these are partially addressed by robustness checks by Berger (1995).

*Estimated relations are reduced form rather than structural*

The first problem is that Equation (1) yields only reduced form estimates, making it difficult to put an economic interpretation on the estimates. Interpreting $\beta_1$ as the causal effect of $k$ on ROE requires the assumption that any association between lagged $k$ and current ROE reflects the effect of $k$ on ROE. As noted above, in econometric terms we treat $k$ as predetermined, which means that future values of ROE do not affect current values of $k$. This removes one potential source of endogeneity in the short run, which is that profits provide a source of capital. Assuming that a bank with ROE higher than expected may retain the additional earnings as capital, higher ROE may result in higher $k$ in the future. However, it is difficult to see how future ROE could affect current $k$, so theoretically at least the equation is protected from this particular concern.

In the long run, it is possible that the level of profitability may affect a bank’s optimal capital ratio. For example, a more profitable bank may expect to have higher ROE in the future and therefore chooses to hold a lower precautionary buffer, knowing it will be able to rely on internal funds to ensure it meets the desired levels in future (Milne and Whalley, 2002). Also, profitability plays an important role in the trade-off theory of capital structure, since the tax deductibility of debt is only relevant for profitable banks which expect to be taxed on their profits. Hence, highly profitable banks may tend to hold lower $k$s as a result. These long-run effects are different from the short run effect of capital described above, but could still be captured by controlling for lagged values of ROE which capture the bank’s expectation about its future ROE.

It is also possible that $\beta_1$ could reflect a bank adopting an expansionary strategy in which it seeks to increase its market share, by reducing margins in order to expand its loan book. Since the bank has cut its margins it is likely that the bank will experience a decline in ROE, at least during the period of transition. However, $k$ may also fall since the bank’s assets are increasing beyond its ability to retain or raise capital. This could plausibly drive correlation
between ROE and lagged $k$ since the banks may raise debt in advance of expanding its loan book, or it may take time for the lower interest rate on new loans to be reflected in average margins across the whole portfolio. This strategy could therefore be reflected in a positive $\beta_1$.

*Lack of differentiation among banks*

The sample of US banks is huge, with around 15,000 banks in 1977 falling to 7000 in 2010. This includes many small banks, some of them regionally focused, as well as very large national banks. The banking sector is highly concentrated; banks in the top third of assets account for over 91% of total bank assets in 1977 and 96% in 2010. Those in the bottom third are only 1-2% of total assets. This makes economic interpretation difficult. For example, an interesting implication of the results is that under distress, banks have substantial incentive to raise their capital ratios and we may therefore expect responses such as scaling back of lending and increased lending margins. The impact this will have on the economy as a whole clearly depends on whether the estimated effects above are different for small and large banks. The effects on aggregate intermediation are likely to be very small if the results are primarily driven by small banks.

A further reason why disaggregation is important is that that there are likely to be differences between banks in terms of their actual and optimal capital ratios. If shocks are to some extent correlated within a group of similar banks, then those banks are likely to be above or below their optimal capital ratios at the same time. Banks which share key characteristics such as size and portfolio risk are also likely to have similar optimal capital ratios. This could complicate interpretation of the estimated coefficients, since it means that whether a particular bank’s $k$ is low or high depends on that bank’s optimal $k$. This would make it difficult to capture non-linearities as described above; for example, if banks were split into “low capital” and “high capital” groups using their actual $k$s, this would not successfully sort them into above and below optimum $k$ since banks have varying optimal $k$s. Therefore, for estimating non-linearities, it seems to make sense to look at disaggregated sub-samples of banks.

*Non-linear effects of capital*

According to the theory put forward by Berger (1995), banks which are below their optimal capital ratio will exhibit a positive $\beta_1$ since they may increase ROE by raising $k$, and those
which are above their optimal capital ratio will exhibit a negative relationship for the same reason. However, the specification in Equation (1) does not allow any differentiation in the cross-section between banks with low capital and those with high capital. Instead, the interpretation of the coefficients relies on the assumption, based on background information, that banks on average were below their optimal capital ratios in the late 1980s and above them in the early 1990s. Even if this is true, since the evolution of $k$ involves a degree of randomness due to unexpected shocks in profitability or market opportunities, we would still expect at least some banks on either side of the optimum in any given period. In other words, while banks on average may be below their optimal $k$ in the late 1980s, there will still be a few banks with very high $ks$ which will exhibit a negative relationship. Therefore, it seems sensible to allow for non-linearity in the $k$-ROE relationship.

An additional reason why this might be desirable is that the existence of such non-linearities provides stronger evidence that we are observing the effects of profitability on $k$, rather than the other way around, as discussed above. The reason is that there is no clear explanation for why the effect of ROE on $k$ would produce such non-linearities. In the example given above, the bank’s expansionary business strategy leads both ROE and $k$ to fall, but this co-movement would be the same for low or high capital banks. In contrast, the “expected bankruptcy costs” hypothesis put forward by Berger (1995) would produce a more positive relationship low capital banks than for high capital banks. In this way the cross-section could be used to increase the robustness of the analysis.

In Table 7 of Berger (1995) he reports estimates of the effect of $k$ on ROE for three categories of $k$ and three categories of risk (defined as risk-weighted assets over total assets). Consistent with the above discussion, the positive effect is stronger for banks with low $k$ and high risk.

**Issues with ROE as the dependent variable**

While ROE is a commonly used measure of a company’s performance, it is flawed in a number of ways. First, it is much more important from a corporate finance viewpoint to capture the value created by different decisions by bank managers. Consequently, the ideal measure of performance would be the total market value of the bank (or, more precisely, the market values of debt and equity). The problem is that the market value of the bank’s assets is not readily measurable, and hence studies tend to focus on book values of equity or assets which reflect historical accounting measures. Mehran and Thakor (2011) provide a good discussion of this issue and provide analysis based on bank M&A, since during M&A there is
an estimate of the value of assets and hence the goodwill captured by the M&A. In our dataset the market value of assets is not available but still it is not clear why it makes sense to focus on ROE rather than return on assets (ROA). Since our aim is to see how the profit margin generated by a given asset depends on how the asset is funded, it would make more sense to look at the ROA.

Some preliminary results are provided in Table 3 of Berger (1995) in which the components of earnings (revenues, interest expense and operating expense) are used in place of ROE in equation (1), expressed both as a ratio of equity and of assets. These results are consistent with the hypothesis that higher $k$ reduces the cost of debt.

4 Towards an improved model of capital and profitability

In this section we present initial results from a model that attempts to deal with some of the issues with the specification in Berger (1995) discussed above. We do this by introducing the notion of a long-run target capital ratio for each bank in each time period. Our main hypothesis is that banks’ optimal capital ratios change over the cycle, and therefore banks can increase their profitability by adjusting upwards or downwards towards the optimal capital ratio. As argued above, this is primarily a short-run phenomenon since, once banks achieve their optimal capital ratios, there is an approximately zero relationship between capital and profitability. In the long run, the optimal capital ratio itself may be affected by the level of profitability. If this hypothesis is correct, then short-run deviations of capital from the optimal capital ratio will be correlated with profitability, after controlling for the long-run co-movement of profits and capital. Furthermore, if the hypothesis of an optimal capital ratio driving the relationship between capital and profitability is correct, negative deviations (deficits) would be positively correlated with profitability, and positive deviations (surpluses) would be negatively correlated with profitability. Intuitively, a bank which is below its optimal capital ratio can improve profitability by increasing capital, whereas a bank which is above its optimal capital ratio can improve profitability by reducing capital.

One important issue is that while our hypothesis above is expressed in terms of the optimal capital ratio, in the analysis below we refer instead to the target capital ratio. The reason is
that the optimal capital ratio is unobserved, perhaps even by the bank itself. Assuming that banks on average tend to move towards their optimal capital ratios over time (since they have profit-maximising incentives to do so) the optimal capital ratio can be approximated by the long-run target capital ratio. In our model, the long-run target capital ratio is the average capital ratio for a particular bank in a given time window, adjusting for the bank’s portfolio risk and the competitive position of the markets in which it operates. However, this provides an imperfect measure of the optimal capital ratio, for two reasons. First, there may be unobserved bank-specific factors which we do not capture, such as changes in regulatory pressure behind the scenes. Second, a bank with adjustment costs will tend to lag behind its optimal capital ratio. For example, if a bank is on average below its optimal capital ratio for all or most of a time period, our estimate of the target capital ratio will be an underestimate of the target capital ratio in that period, although such discrepancies will tend to average out to zero over the long run. 

We address two further issues discussed above. We increase the amount of heterogeneity in the model by estimating the whole model separately for different groups of banks defined in terms of size and portfolio risk. This allows for more closely fitted estimates of the target capital ratio relevant to each group of banks since it allows for non-linearities in the estimated model over the dimensions of size and risk. It also allows us to see how relevant our hypothesis is for different groups of banks; for example, it seems likely that expected bankruptcy costs would be more relevant for banks with a higher degree of portfolio risk.

We also switch from ROE to ROA as the measure of profitability. We do not expect the results to be very different on the two measures, since \( \text{ROA} = \text{ROE} \times k \).

### 4.1 Specification

Our model consists of three stages. In the first stage, we rank banks in terms of their size and portfolio risk in each year and distribute them across nine size-risk groups. Second, we estimate a long run target capital ratio on annual data for each size-risk group \( \kappa \) and window

---

8 A possible robustness check would be to alter the weightings given to different periods, so that the target capital ratio reflects the future more than the past. Of course, the disadvantage of this is that future data will also reflect information unavailable to the bank at time \( t \), making it a rather noisy estimate of the target capital ratio at \( t \).

9 For a similar example of a panel approach exploiting cross-sectional heterogeneity (but applied to countries instead of banks), see Brun-Aguerre et al (2011).
τ. Third, the target capital ratio is included in a model of return on assets, similar to the specification in Equation (1). We also include in stage 3 measures of the deviation of a bank’s capital ratio from the target capital ratio. Since our hypothesis of interest concerns the short-run relationship we estimate stage 3 using quarterly data.

**Stage 1: Dividing banks into groups by size and risk**

In the first stage, we split banks into nine groups denoted κ according to their size and level of portfolio risk. We use the same five measures of portfolio risk used in Equation (1) above, namely the ratio of commercial loans to total assets (C&I), the standard deviation of return on equity (SDROE), the ratio on net charge-offs to total assets (CHARGE), the ratio of non-performing loans to total loans (NPL), and the ratio of risk-weighted assets to total assets (RW). We aggregate these variables into a single risk index, adopting the simple strategy of assigning each measure equal weight. For each quarter, we rank banks according to each risk measure and calculate the fraction of the total number of banks to give an index between 0 and 1. The risk indices are summed (using all those available in a particular year), and we then rescale the resulting measure to calculate a single risk index $RISK$ taking values between 0 and 1 in each quarter. We follow the same approach for size, ranking banks in each quarter according to their total assets and calculating an index $SIZE$ taking values between 0 and 1.

We use the risk and size indices to assign banks to the nine groups. Banks in the bottom 1/3 of the risk index and the bottom 1/3 of the size index are assigned to the small-low risk group; banks in the middle 1/3 of the risk index and the bottom 1/3 of the size index are assigned to the small-medium risk group, and so on. The number of banks and the mean values of $RISK$ and $SIZE$ are shown for each group in Table 4.1, for three selected quarters at the beginning, middle and end of our sample. The resulting groups are uneven in size due to changes in the number of banks over time and correlation between risk and size. Big banks are more likely to be high risk than small banks, so the big-high risk group contains more banks than big-low risk; the opposite is true for small banks, which tend to be low risk.

<table>
<thead>
<tr>
<th>κ</th>
<th>1977Q1</th>
<th>1995Q1</th>
<th>2010Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means of:</td>
<td>N</td>
<td>Means of:</td>
</tr>
<tr>
<td></td>
<td>RISK</td>
<td>SIZE</td>
<td>RISK</td>
</tr>
<tr>
<td>1</td>
<td>Small, low risk</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>Small, medium risk</td>
<td>0.49</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>Small, high risk</td>
<td>0.83</td>
<td>0.19</td>
</tr>
</tbody>
</table>

**Table 4.1: Descriptive statistics for size and risk groups**
Stage 2: Estimating target capital ratios

In the second stage, we estimate target capital ratios for each bank and time period. We do this separately for each group $\kappa = 1$ to $9$, using annual data (using only the fourth quarter in each year as above). This is done for 5-year rolling windows $\tau = 1$ to $30$ as above. Our estimation equation for each group $\kappa$ and window $\tau$ is:

$$k_{it} = a_{1\kappa}^1 + a_{2\kappa}^2 k_{it-1} + b_{\kappa}^1' X_{it-1} + u_{i,\kappa,\tau} + \epsilon_{it} \tag{2}$$

Note that this specification takes the form of a partial adjustment model with a long-run equation as follows:

$$k_{it}^* = A_{\kappa} + B_{\kappa}^1 X_{it} + U_{i,\kappa,\tau} \tag{3}$$

Where $B_{\kappa} = b_{\kappa}/(1 - a_{2\kappa}^2)$ and the time-invariant long-run effect for each bank is given by $A_{\kappa} + U_{i,\kappa,\tau} = (a_{1\kappa}^1 + u_{i,\kappa,\tau})/(1 - a_{2\kappa}^2)$. The control variables in $X_{it}$ include the controls for portfolio risk, market structure and operating efficiency set out in 3.1, namely SDROE, CHARGE, NPL, C&I, RW, BANKSHARE, MSA, HERF, GROWTH and AC. We also include lagged ROA to control for the long-run effect of profitability on target capital ratios. Since we have split the sample into groups according to size, the use of size dummies is no longer the most appropriate treatment of size, and instead we control for the effects of size within size-risk groups $\kappa$ by including the size index SIZE described above. Due to the presence of a lagged dependent variable, we estimate using two-stage system GMM with corrected standard errors, as described in 3.1 above. Both lagged $k$ and ROA are treated as endogenous and instrumented using lags of themselves and the other explanatory variables.
We calculate from (2) and (3) a long-run target capital ratio for each bank \( k_{tit}^{*} \). This is bank- and time-specific, but it also depends on which sample window \( \tau \) is used. Equation (3) is estimated using rolling windows which overlap with each other, and so in any given year we have up to 5 estimates of the target capital ratio \( k_{tit}^{*} \) available for each bank, based on years in the past and future.\(^{10}\) We calculate our final estimate of the target capital ratio \( k_{tit}^{*} \) by calculating the average of all of the estimates that are available in a particular year. This is equivalent to a weighted average where greater weight is given to estimates based on closer time periods.\(^{11}\)

**Stage 3: Estimating the determinants of the return on assets**

The final stage of our analysis is to analyse the effects of the capital ratio on the return on assets, using a modified version of Equation (1), incorporating the long run target capital ratio from (3). First, we merge the annual \( k_{it}^{*} \) from Equation (3) with the quarterly data, applying linear interpolation to obtain \( k_{it}^{*} \) for quarters 1-3 (note that from this point and for the rest of the paper \( t \) refers to quarters rather than years). We then derive variables to capture surpluses and deficits of \( k_{it} \) relative to \( k_{it}^{*} \), calculated by multiplying through by dummy variables for banks in surplus and in deficit:

\[
    k_{it}^{Surplus} = (k_{it} - k_{it}^{*}) \cdot [k_{it} > k_{it}^{*}] \\
    k_{it}^{Deficit} = (k_{it} - k_{it}^{*}) \cdot [k_{it} < k_{it}^{*}]
\]

Once we have calculated \( k_{it}^{Surplus} \) and \( k_{it}^{Deficit} \) for each bank-year observation, we winsorize them at the 1% and 99% levels in each year, in order to reduce the influence of extreme observations.

---

\(^{10}\) Using future data to parameterise the target capital ratio may seem to be counter-intuitive, but we would argue that it is desirable since banks tend to experience lags in adjusting to their target capital ratios. On the one hand, the capital ratio at \( t+1 \) tells us something about the target capital ratio at \( t \), but on the other hand this has the disadvantage that the capital ratio at \( t+1 \) will also reflect information about future conditions that the bank is unaware of at time \( t \).

\(^{11}\) For example, for the year 1985 estimates based on the equations from 1981-1985 to 1985-89 are relevant. 1981 gets included in only one of the estimates used to calculate the target capital ratio for 1985, 1982 is included in two, 1983 in three, and so on. The greatest weight is given to the year 1985 itself which is included in all 5 sets of estimates. As a percentage of the years used as inputs to the final sample, 1985 gets a 20% weight, 1984 and 1986 get 16%, 1983 and 1987 get 12%, 1982 and 1988 get 8% and 1981 and 1989 get 4%.
\( k_{it}^*, k_{it}^{\text{Surplus}} \) and \( k_{it}^{\text{Deficit}} \) are included in a model of ROA, which is estimated on quarterly data, once again separately for each size-risk group \( \kappa \) and window \( \tau \):

\[
ROA_{it} = \alpha_{\kappa \tau} + \theta_t + u_i + \sum_{j=1}^{J} \delta_{\kappa \tau \tau - j} ROA_{t - j} + \sum_{j=1}^{J} \pi_{\kappa \tau j}^1 k_{it}^* + \sum_{j=1}^{J} \pi_{\kappa \tau j}^2 k_{it - j}^{\text{Deficit}} + \sum_{j=1}^{J} \pi_{\kappa \tau j}^3 k_{it - j}^{\text{Surplus}} + \beta_{\kappa \tau} Z_{it} + \varepsilon_{it}
\]

In this equation \( \pi^2 \) gives the long-run relationship, i.e. where \( k = k^* \), and \( \pi^2 \) and \( \pi^3 \) give the short-run effects, where \( k^* \) is fixed. If our hypothesis about the effect of capital on profitability is correct, that banks tend to target their optimal capital ratios where profitability is maximised, then once the long-run effects of capital have been removed by controlling for \( k_{it}^* \), \( k_{it - 1}^{\text{Deficit}} \) should be negatively correlated with \( ROA_{it} \) (\( \pi^3 < 0 \)) and \( k_{it - 1}^{\text{Deficit}} \) should be positively correlated with \( ROA_{it} \) (\( \pi^2 > 0 \)). We again estimate Equation (4) using 5-year rolling windows. There are 20 time periods in each 5-year window, and 29 years in total (the first year, 1977, is lost due to the inclusion of (annual) lagged variables in Equation (2). As before, equation (4) is estimated using two-step system GMM, instrumenting lags of ROA. \( J \) is set to 4 to capture a full year of lags.

### 4.2 Results

We show the detailed results from estimating our target capital ratio Equation (2) in Table 4.2, Table 4.3 and Table 4.4 for small banks, medium sized banks and big banks respectively. Considering the large number of regression results, we only show the long-run coefficients from Equation (3), and rather than showing results for all 30 windows we only show those for selected windows spanning the whole range of our sample period, 1977-1980, 1981-85, 1986-90, 1991-95, 1996-2000, 2001-05 and 2006-10. Among those variables with most consistent explanatory value are ROA which is positively correlated with \( k \) and SIZE which is negative, suggesting effects of size within the size groups. BANKSHARE is positively correlated with \( k \), suggesting that banks with a dominant local market position tend to have
higher capital ratios, but other market structure variables have mixed sign and significance. Of the portfolio risk measures, CHARGE is positive and significant for medium and high risk banks, whereas C&I, SDROE and RW are negative and NPL is mixed in sign and significance. The latter finding may indicate that a riskier business profile reduces the bank’s charter value and therefore reduce incentives for holding higher capital, or it could simply be that banks tend to target a given probability of default, so more risk averse banks tend to have higher portfolio risk and higher capital (Jokipii and Milne 2011). Finally, OPEFF has mixed sign and significance. The diagnostic tests produce rather mixed results. The Arellano-Bond test for auto-correlation in differenced residuals indicate that there is autocorrelation of order 1 but, for most samples not in order 2 which validates the use of lagged levels of k and ROE as instruments. However, the Sargan and Hansen tests generally reject the null hypothesis of exogeneity of the instrument set. This may be because of the use of portfolio risk variables as explanatory variables, which are potentially endogenous.

Having estimated the long-run determinants of the k, we proceed to calculate the long-run target capital ratio using Equation (3). In Figure 4.1, we show the median k and target k in each quarter (dashed and full lines respectively), separately for each group. Consistent with the general rise in k shown in Figure 2.1 and Figure 2.2 above, the charts show that the k rose substantially over our sample period across every level of portfolio risk and size. The k and the target k are shown to move closely together over the sample period and both exhibit the rise between the 1990s and the 2000s. We also show on the right axis the buffer used to calculate $k_{it}^{Surplus}$ and $k_{it}^{Deficit}$, in bands between the 10th and 90th percentiles (light gray) and between the 25th and 75th percentiles (dark grey). For most groups and quarters, these lie between +/- 2 percentage points, although they can be seen to rise and fall over time as banks ks move around the target k over time. Most banks are below their target capital ratios in the mid-1990s, consistent with the substantial rise in ks that was observed at that time. However, k are not below target k in the mid- to late- 1980s, counter to the argument in Berger (1995) that optimal ks were rising in this period; in fact for the high risk groups ks are generally

---

12 The Im-Shin-Pesaran panel unit root test was used to test the null hypothesis that all of the panel units have unit roots and this was rejected, consistent with the economic rationale that k should be stationary. However, some of the regressions of equation (2) exhibit a high degree of autocorrelation of k, which may indicate that the use of the partial adjustment model is invalid, since high values of $a_{kT}^2$ produce implausibly large estimates of the long-run coefficients (this can be seen, for example, in the estimates for group 1 in 1977-80 in Table 4.2 and group 7 in 1991-95 and 1996-2000 in Table 4.4). This suggests that a unit root may be relevant in some subsamples, and I adopt the practical step of replacing the estimated target capital ratio with the 5-year moving average of k for those windows with $a_{kT}^2 > 0.85$ (approximately the value above which $a_{kT}^2$ is not significantly different from one). This affects 19 out of 270 group-window pairs.
above their target levels in this period. This could reflect that the target $k$s are lagging indicators of banks’ optimal capital ratios and therefore may not capture the shift in unobserved optimal capital ratios that took place.
Table 4.2: Long-run target capital structure, groups 1-3 (small banks)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SIZE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANKSHARE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HERF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROWTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHARGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C&amp;I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDROE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPEFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes: The table shows the long run coefficients derived from estimation of reduced form Equation (2) for three groups: 1 (small, low risk), 2 (small, medium risk) and 3 (small, high risk). Table shows 7 windows per group selected from the 30 that are estimated. The capital asset ratio (k) is regressed on one lag of itself and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of the bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C&amp;I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL), the ratio of risk-weighted assets to total assets (RW), our size index based on total assets (SIZE) and the ratio of operating costs to total assets (AC). Bank fixed effects are included (but not time effects). Estimated using two-step system GMM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3: Long-run target capital structure, groups 4-6 (medium-sized banks)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SIZE</td>
<td>-0.003* -0.019*** -0.026*** -0.018*** -0.017*** -0.035*** -0.026***</td>
<td>0.013*** -0.022*** -0.021*** -0.015*** -0.019*** -0.019*** -0.018***</td>
<td>0.019*** -0.015*** -0.003 -0.013*** -0.013*** -0.016*** -0.011***</td>
</tr>
<tr>
<td>ROA</td>
<td>-5.586 0.999*** 0.561* 0.276 0.671* 1.491*** 2.703***</td>
<td>0.525*** 0.316*** 0.45*** 0.393 1.032*** 0.652*** 0.713***</td>
<td>0.468*** 0.531*** 0.395*** 0.692*** 0.448*** 0.51*** 0.818***</td>
</tr>
<tr>
<td>BANKSHARE</td>
<td>-0.016 0.008* 0.011** 0.022*** 0.058*** 0.061*** 0.022***</td>
<td>0.008*** 0.016*** 0.015*** 0.011* 0.027*** 0.005</td>
<td>0.014*** 0.009*** 0.011*** 0.016*** 0.035*** 0.014* 0.025***</td>
</tr>
<tr>
<td>HERF</td>
<td>0.069 0.008 0.004 -0.015*** -0.031 -0.026 -0.018</td>
<td>0.003 -0.002 0.001 0.004 -0.011 -0.021*** 0.001</td>
<td>-0.002 0.007*** 0.009* -0.005 -0.018** -0.005 0.003</td>
</tr>
<tr>
<td>GROWTH</td>
<td>-0.029 -0.001 &gt;0.001*** &gt;0.001 &gt;0.005 &gt;0.002 &gt;0.002</td>
<td>-0.003* -0.002 &gt;0.001*** &gt;0.001 &gt;0.001 &gt;0.001 &gt;0.001</td>
<td>&gt;0.001 &gt;0.001 &gt;0.001 &gt;0.001 &gt;0.001 &gt;0.001 &gt;0.001</td>
</tr>
<tr>
<td>MSA</td>
<td>-0.006 -0.003*** -0.005*** -0.004* 0.007* 0.007* 0.012***</td>
<td>-0.001 &gt;0.002 0.002* &gt;0.001 &gt;0.001 0.002 0.005***</td>
<td>0.003*** 0.001* &gt;0.001 0.003* 0.006*** 0.003 0.006***</td>
</tr>
<tr>
<td>CHARGE</td>
<td>1.919 0.012 0.02 -0.369 -1.161 1.138 -2.003</td>
<td>-0.018 0.011 1.811*** 0.282 1.128*** 0.703*** 1.23***</td>
<td>0.045*** 0.004 0.178*** 0.586*** 0.652*** 0.681*** 0.524***</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>0.08 -0.027*** -0.045*** -0.071*** -0.126*** -0.063*** -0.092***</td>
<td>0.01* -0.016*** -0.005 -0.018 -0.038*** -0.007 -0.024***</td>
<td>0.07 -0.017*** -0.018*** -0.005*** -0.013 -0.009 -0.001</td>
</tr>
<tr>
<td>NPL</td>
<td>- - 0.052 0.036 0.346*** 0.314*** 0.342***</td>
<td>- - 0.047*** 0.118*** 0.312*** 0.349*** 0.111***</td>
<td>- - -0.231*** -0.068 0.074 0.133*** -0.152***</td>
</tr>
<tr>
<td>OPEFF</td>
<td>- - (0.04) (0.08) (0.14) (0.134) (0.157)</td>
<td>- - (0.023) (0.048) (0.093) (0.065) (0.063)</td>
<td>- - (0.023) (0.042) (0.05) (0.043) (0.035)</td>
</tr>
<tr>
<td>RW</td>
<td>- - - - - - -</td>
<td>- - (0.097) (0.145) (0.153) (0.134) (0.088)</td>
<td>- - (0.098) (0.135) (0.164) (0.069) (0.122)</td>
</tr>
</tbody>
</table>

Notes: The table shows the long-run coefficients derived from estimation of reduced form Equation (2) for three groups: 1 (small, low risk), 2 (small, medium risk) and 3 (small, high risk). Table shows 7 windows per group selected from the 30 that are estimated. The capital asset ratio (k) is regressed on one lag of itself and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of a bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C&I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL), the ratio of risk-weighted assets to total assets (RW), our size index based on total assets (SIZE) and the ratio of operating costs to total assets (AC). Bank fixed effects are included (but not time effects). Estimated using two-step system GMM.
### Table 4.4: Long-run target capital structure, groups 7-9 (big banks)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SIZE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANKSHARE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HERF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROWTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHARGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDROE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPEFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The table shows the long run coefficients derived from estimation of reduced form Equation (2) for three groups: 1 (small, low risk), 2 (small, medium risk) and 3 (small, high risk). Table shows 7 windows per group selected from the 30 that are estimated. The capital asset ratio \( k \) is regressed on one lag of itself and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of a bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C/I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL), the ratio of risk-weighted assets to total assets (RW), our size index based on total assets (SIZE) and the ratio of operating costs to total assets (AC). Bank fixed effects are included (but not time effects). Estimated using two-step system GMM.
Notes: Charts show the medians of the actual $k$ and $k^*$ (left axis), estimated using equations (2) and (4), full details of which are given in the notes to Table 4.2, Table 4.3 and Table 4.4. We also show the buffer of $k$ relative to $k^*$ at the 10th, 25th, 75th and 90th percentiles (right axis).
Next we proceed to our estimation of equation (4) which shows the relationship between ROA and the three capital variables $k_{it}^*, k_{it}^{Surplus}$ and $k_{it}^{Deficit}$, as well as the control variables for portfolio risk and market structure. As a reminder, the target capital ratio was estimated using equation (2) and has been interpolated from annual data to quarterly data in order to run the ROA regression (4). For brevity, we show only the total coefficients on the capital variables in (4) in Figure 4.2. Overall, the long-run target capital ratio is positively correlated with lagged ROA; this is not surprising since in the majority of the target capital regression the long-run coefficient was also positive. $k_{it}^{Deficit}$ is mostly positive as expected, but is also negative for a large number of banks, whereas $k_{it}^{Surplus}$ is consistently negative and strongly significant.

We first examine the sample period 1983-89 used by Berger (1995). If the “expected bankruptcy costs” hypothesis in that paper is correct, then we would expect banks with a deficit of capital relative to target to exhibit $\pi^2 < 0$. Inspection of Figure 4.2 shows that this is observed, but only for the highest risk groups (3, 6, 9), which is consistent with the idea that the riskiest banks would be under pressure from the market to increase their capital ratios under stressed market conditions. For the other groups of banks, there is no evidence that banks with low capital relative to their long-run targets are able to increase ROA by raising $k$ any more than other banks, which calls into question whether the “expected bankruptcy costs” hypothesis is relevant for these banks. we also find that during this period, banks with a capital surplus relative to the long-run target are able to increase profitability by reducing capital ratios ($\pi^3 < 0$). This indicates that banks with relatively high capital ratios are able to increase ROA by reducing $k$, at least in the short run. This validates our more heterogeneous specification, since controlling only for $k$ would not reveal this interesting finding.

The next period of our analysis is the 1990s and early 2000s. As a reminder, our analysis above showed that this is a period of rapidly rising capital ratios and strong profitability. For small banks, a deficit of capital is generally associated with a positive $k$-ROA relationship though the results are fairly weak and only observed for the low and medium risk groups. A surplus of capital has a coefficient that is strong and consistently negative for all three groups of small banks, suggesting that banks with high capital ratios were able to raise ROA by reducing $k$. For both big and medium sized banks during this period there is a positive though weak relationship (generally not significantly different from zero) for medium and high risk banks with a capital deficit, and again a capital surplus is linked to a strongly negative and statistically significant relationship. For all nine groups, the coefficient on the long-run target $k$, $\pi^1$, is positive and significant. These results offer some insight into the reasons for the increase in capital ratios during this period. While capital ratios are rising strongly across the sample period, low capital banks do not improve their profitability by increasing
their capital ratios, at least in the short run, and this is true for low and high risk banks. Thus the results do not support the contention of Flannery and Rangan (2008) that weakening of implicit government guarantees increased the capital demanded by investors in this period. The long-run co-movement of $k$ and ROA is strongly positive, supporting the idea that high profitability was driving higher capital ratios during this period, rather than the other way around.

Finally, we examine the results for the recent crisis period, 2007-2010. As noted above, in this period banks suffered substantial losses and have received unprecedented levels of official support. Capital ratios have also come under pressure as banks seek to show their balance sheets are solvent. However, our findings do not support the idea that banks with low capital relative to long-run target levels were able to increase their ROA, and this was true as much for high risk as for low risk banks (with the exception of small high risk banks for whom the coefficient $\pi^2$ is significantly positive). The long-run target capital was positively linked to capital but to nothing like the same extent that was observed in the 1980s crisis. For most groups including the high risk groups, it remains true in this period that banks with high capital ratios can improve profitability by reducing their capital ratios. These results are slightly surprising coming from a period of huge turmoil in the financial sector. One possible explanation of these results is that official support and, indeed, injections of capital by the government directly into troubled banks, have softened the effects of market discipline in this period.

One possible interpretation of the weak results for capital deficits, relative to capital surpluses, is the role of regulatory capital requirements. If banks set their long-run target capital ratio as a buffer over the capital requirements, then this implies that falling below the target would be costly for the bank due to the risk of supervisory intervention (see Milne and Whalley 2002 and other references cited in 2.1 above). If investors know and care about these supervisory costs, then they would punish banks that have capital deficits, suggesting lower profitability. Hence, the result is a modified version of the trade off theory where regulation causes value to be upwards sloping in capital below the optimal capital ratio. However, if investors do not price in the costs of supervisory intervention, for example because they are unaware of supervisory activities (Berger et al 2000) or because they believe they will be shielded from losses by government assistance, then the result (assuming capital requirements are high enough to be binding on the bank) is that even below the long-run target capital ratio profitability may be negatively correlated with capital, since investors see the capital requirements as value-destroying. This may suggest why deficits of capital are often negatively or non-correlated with capital ratios.
Figure 4.2: Coefficients on capital variables in regression of ROA

Notes: The charts show the total coefficients on $k_{it}^1$, $k_{it}^{Surplus}$, $k_{it}^{Deficit}$ from estimation of (4) on quarterly data and 5-year estimation windows. The return on assets (ROA) is regressed on four lags of itself, four lags of the capital variables, and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of a bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C&I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL, after 1984), the ratio of risk-weighted assets to total assets (RW, after 1996), the size rank index based on total assets (SIZE) and the ratio of operating costs to total assets (AC, after 1984). Bank and time fixed effects are included. Estimated using two-step system GMM. Bands show two standard errors (~95%).
Figure 4.2: Coefficients on capital variables in regression of ROA (continued)

Notes: The charts show the total coefficients on $k_{it}^1$ ($\pi_1$), $k_{it}^3$ ($\pi_3$) and $k_{it}^2$ ($\pi_2$) from estimation of (4) on quarterly data and 5-year estimation windows. The return on assets (ROA) is regressed on four lags of itself, four lags of the capital variables, and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of a bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C&I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL, after 1984), the ratio of risk-weighted assets to total assets (RW, after 1996), the size rank index based on total assets (SIZE) and the ratio of operating costs to total assets (AC, after 1984). Bank and time fixed effects are included. Estimated using two-step system GMM. Bands show two standard errors (~95%).
Figure 4.2: Coefficients on capital variables in regression of ROA (continued)

Notes: The charts show the total coefficients on $k_{it}^{\pi_1}$, $k_{it}^{\text{Surplus}}$, and $k_{it}^{\text{Deficit}}$ from estimation of (4) on quarterly data and 5-year estimation windows. The return on assets (ROA) is regressed on four lags of itself, four lags of the capital variables, and lags of the following control variables: the bank’s share of local deposit markets (BANKSHARE), the Herfindahl index for the bank’s local deposit markets (HERF), annual growth of a bank’s local deposit markets (GROWTH), the proportion of a bank’s deposits in urban markets (MSA), the ratio of charge-offs to total assets (CHARGE), the ratio of commercial and industrial loans to total assets (C&I), the standard deviation of ROE (SDROE), the ratio of non-performing loans to total assets (NPL, after 1984), the ratio of risk-weighted assets to total assets (RW, after 1996), the size rank index based on total assets (SIZE) and the ratio of operating costs to total assets (AC, after 1984). Bank and time fixed effects are included. Estimated using two-step system GMM. Bands show two standard errors (~95%).
5 Conclusions

In this paper, we have revisited the effect of capital on profitability in US banks over the period 1977-2010 in order to assess whether the relationship varies across banks according to whether they have high or low capital, and whether the relationship is time-varying as a result. We have carried out the analysis in two parts. First, we extend the results of Berger (1995), adding data from 1993-2010 in order to carry out a more systematic analysis of the relationship over an extended time period than has been conducted previously. Consistent with that study, we find a positive relationship in the 1980s during the so-called savings and loan crisis. The same finding is true, though to a lesser extent, during the recent GFC, since the effect of capital on ROE is close to zero or marginally positive. However, we have also discussed several disadvantages with the specification used by Berger (1995), namely that the use of a reduced form specification may fail to disentangle the effects of capital on profitability from the long-run effects of profitability on capital. Indeed, there are a number of prominent theories of why profitability would drive capital in the long run.

Therefore, in the second part of the analysis we have presented results from an improved version of the model in which we include not only the effect of the long-run target capital ratios but also the surplus or deficit of capital ratio from the long-run target. Exploiting the cross-section of the sample in this way allows a more robust test of the hypothesis that capital affects profitability differently depending on whether banks are at above or below their optimal capital ratios. The specification allows considerable heterogeneity in the specification of target capital ratios and in the model of profitability in order to accurately capture whether banks are close to their optimal capital ratios. We find that the while the long-run relationship between capital ratios and ROA is consistently positive, there are asymmetric effects of deviations from the long-run target capital ratio; deficits tend to be positively associated with future ROA whereas surpluses tend to be strongly negatively associated with future ROA. This is consistent with the trade-off theory of capital structure and suggests that under stressed conditions, banks may be able to improve their profitability by increasing capital ratios.

The findings largely support the conclusions of Berger (1995) that banks in the late 1980s savings and loan crisis were able to improve their profitability by increasing their capital ratios. We find that this is true predominantly for banks with high portfolio risk. In the
1990s and early 2000s, the asymmetry is not quite so apparent and there is little evidence that low capital banks were able to improve profitability by raising capital ratios. This stands in contrast to the observation that capital ratios were rising rapidly over this time period, often well in excess of capital requirements (Flannery and Rangan 2008, Berger et al 2008). While studies such as Flannery and Rangan (2008) have argued that increased market discipline played a role in the rise in capital ratios, our findings suggest instead that this is not the case; indeed banks with high capital ratios relative to their long-run targets could improve profitability by reducing capital ratios rather than by increasing them, and there is no evidence that banks with low capital ratios relative to target could increase profitability by increasing capital ratio. With respect to the recent crisis period, there is a stark contrast between the findings from the model based on Berger (1995) and the improved model, since while the basic model finds that increasing capital drives (Granger causes) an increase in ROA in the 2008-10 period, the improved model does not support the conclusion that low capital banks were able to improve ROA by raising capital. Again we find a strong negative relationship for banks with high capital relative to their long-run target capital ratios. These results may reflect the role of official support in softening the effect of market discipline.

These results have important policy implications for the operation of capital requirements. The cost of capital requirements depends on where they are set relative to banks’ own desired level of capital. As we have shown, banks tend to have a desired level of capital above which increases in capital may reduce profitability. Capital requirements, if they are binding, are likely to have a larger effect on banks’ costs the higher above the optimal capital ratio they are set. Since the optimal capital ratio is likely to vary over financial cycles, the costs of capital requirements will vary as well, which is a factor that has not been taken into account in most assessments of the impact of capital requirements. This is particularly relevant for macroprudential policy. This policy operates through a regime of time-varying capital requirements, which rise during booms and fall during busts in order to smooth banks’ credit supply over the cycle. If optimal capital requirements rise during a crisis, it is unlikely that banks will have incentives to reduce their capital ratios if capital requirements are cut. This may limit the effectiveness of counter-cyclical capital requirements, since reductions in the capital requirement will have little effect on banks whose optimal capital ratios meet or exceed the level of the capital requirement. On the other hand, increases in capital requirements during a boom period may raise a bank’s capital ratio above the desired level, consistent with the aims of the policy.
References


