We would like to thank Sid Winter for valuable insights that helped build and then sharpen the theory; Doug Duncan for his support and guidance in the mortgage banking industry; Freek Vermeulen, Julian Birkinshaw, Sumantra Ghoshal, Costas Markides and Margarethe Wiersema for useful comments; and Srikanth Kannan for research assistance. We would also like to thank Will Mitchell and two anonymous SMJ referees for their useful comments. This work was partially funded by National Science Foundation Grant IIS-9733877, the Centre for the Network Economy and the Leverhulme Trust Digital Divide Programme at the London Business School, and was made possible through generous funding by the Mortgage Bankers Association of America. Part of the work was done while the first author was a NATO Science Fellow.
LOSING SIGHT OF THE FOREST FOR THE TREES?
PRODUCTIVE CAPABILITIES AND GAINS FROM TRADE AS DRIVERS OF VERTICAL SCOPE

Abstract

Focusing on proving or disproving Transaction Cost Economics has led to a relative neglect of some key drivers of vertical scope, such as differences in productive capabilities (as opposed to capabilities of governance). We consider how productive capability differences shape vertical scope through gains from trade. Using highly detailed data on the mortgage banking industry, we find productive capabilities to be a key determinant of the make-vs.-buy decision. Our analysis also suggests firms’ attempts to leverage a comparative advantage can also lead to the use of mixed governance modes (both “make” and “buy” in a particular part of the value chain). We conclude that the distribution of productive capabilities along the value chain, catalyzed by transaction costs, ultimately drives vertical scope.

Keywords: Productive Capabilities; Gains from Trade; Transaction Costs; Vertical Integration
What determines vertical scope? In the last three decades, transaction cost economics (TCE) has been the dominant paradigm for understanding why firms choose to “make” rather than “buy” required production inputs. The basic TCE argument is that the hazards of the market lead firms to increase vertical scope, while the absence of such hazards and gains from specialization lead firms to use “the market.” Numerous empirical studies on vertical scope have been produced over the last fifteen years, and in most of them transaction costs (TC) have been shown to be a driver of the make-or-buy choice (see Shelanski and Klein, 1995; Williamson, 1999). However, the pursuit of empirical validation of the TCE framework and related theories from organizational economics may have taken the focus away from understanding the different factors that can come into play in explaining vertical scope.

Recently, TCE-based views of vertical scope have been augmented with a nascent “capability-based view” of integration, which focuses on firms’ relative advantages (Argyres, 1996; Poppo and Zenger, 1998; Schilling and Steensma, 2002; Leiblein and Miller, 2003; Hoetker, 2005). However, empirical work drawing on this approach is more limited, and research has not directly considered the relative explanatory power of different perspectives on vertical scope. Our paper contributes to this research stream by comparing existing views on vertical scope from economics and TCE. Our focus is on the “comparative advantage” perspective (see Jacobides and Winter, 2005), which focuses on the role of particular firms’ productive capabilities (the operational efficiency of a portion of a production process) in explaining scope, as contrasted with capabilities of governance which have been studied more extensively (Dyer, 1996; Argyres and Liebeskind, 1999; Madhok, 2002). Firms with greater productive capabilities in a stage of production will tend to perform this activity internally, and contract with another firm through the “market” where they are deficient. Similarly, firms with a comparative strength in one segment (e.g., upstream) may want to use both their own downstream operations and the “market” to leverage their strength. By explicitly recognizing that using the “market” is really using the capabilities of another firm willing to transact, our analysis brings productive capabilities to the forefront.
Vertical specialization can thus be likened to the analysis of international trade: TC imposes a net “tax” to transactions that happen through the market. This net tax is the difference between internal governance costs (driven by bureaucracy) and market procurement costs (driven by coordination costs, information asymmetries, or opportunism and its antecedents such as asset specificity). This net tax deters but does not prevent specialization. Like in international trade, sufficiently high gains from contracting can often overcome procurement costs and lead firms to specialize vertically. Conversely, if productive capabilities do not vary much across firms vertical specialization will not occur, even if the net TC “tax” is extremely low as there is little gain to be had from contracting. Our perspective is analogous to the resource-based view of diversification (Wernerfelt and Montgomery, 1988). When skills or knowledge create an advantage for only one segment of a value chain, firms will tend to specialize; if these skills apply across multiple value chain segments there is little grounds for specialization at the firm or industry level.

Our approach differs somewhat from recent applications of the resource based view (RBV) which has emphasized capabilities of governance (Winter, 1988; Argyres and Liebeskind, 1999; Leiblein and Miller, 2003), rather than productive capabilities (Demsetz, 1988, Langlois and Foss, 1999; Hoetker, 2005). In addition, as our attention is focused on comparative rather than competitive advantage (as in Hoetker, 2005); issues of sustainability and rarity, central to most RBV analyses (e.g. Combs and Ketchen, 1999; Schilling and Steensma, 2002), do not play as critical a role.

In this paper we make three contributions. First, we articulate a “comparative advantage” view of scope as it applies in this industry, focusing on productive capabilities (for greater detail, see Jacobides and Winter, 2005). Second, we empirically test not only the productive capabilities perspective, but also its relative explanatory power compared to other perspectives on the drivers of scope, notably some of the factors identified by TCE - especially the transaction costs due to information asymmetry and measurement problems. Third, we provide a motivation for “mixed governance”, the concurrent use of own operations and market-based arrangements. We argue and
empirically demonstrate that highly capable firms upstream may be driven to use both their captive
downstream divisions and use other downstream specialists so as to leverage their comparative
advantages. This may explain why such “mixed mode” solutions are empirically much more
prevalent than theory might suggest (Harrigan, 1985; Parmigiani, 2004).

Our empirical analysis is made possible by a unique firm-level panel dataset compiled by the
Mortgage Bankers Association of America for quasi-regulatory purposes. Due to the unique
characteristics of mortgage banking in which products and the value chain segments are clearly
and uniformly defined, we are able to measure transaction risks, calculate the efficiency of a firm
in each segment of the value chain, and define a continuous measure of integration across different
segments. On the basis of these data, we find that while both the TCE and the “productive
capabilities” views of vertical scope are supported, but the productive capabilities view appears to
explain much more variance in scope. This suggests that productive capabilities are a central driver
of scope decisions in this industry, and may be important in other industries with substantial
productivity differences across firms in different segments of the value chain.

THEORY AND HYPOTHESES

Several strands of research have provided insights on what drives the make vs. buy decision.
Rather than providing a comprehensive guide to the drivers of vertical scope, this section
summarizes the motivations to vertically integrate, focusing on transactional explanations and on
the role of resources and capabilities.

Institutional and Information Economics: Measurement and Transaction Costs

Perhaps the most studied set of determinants of vertical scope comes from institutional
economics, especially TCE. TCE posits firms will produce intermediate products internally rather
than purchase them on the market the greater the hazards of market exchange. Such hazards arise
due to either to the ex ante problems of information misrepresentation for any given exchange, or
to the risks of asset-specific investment, the value of which may be expropriated ex post by an
opportunistic party in the future. Ex-ante information asymmetry can create situations where market transactions can be costly or impossible due to fear of strategic misrepresentation described by Akerlof (1970) as the “lemons problem”. This information asymmetry can arise from the limited ability to measure or assess intermediate goods (Barzel, 1982) or the lack of a standardized “grammar and syntax” (Argyres, 1999) to describe such goods (Baldwin and Clark, 2003; Jacobides, 2005). Integrated firms, in contrast, have advantages of lower measurement costs or a greater ability to utilize subjective incentives that do not rely on precise measurement (Masten et al., 1991; Jacobides and Croson, 2001). In our setting, firms that originate mortgage loans often have better information about the risk characteristics of the borrower than firms that purchase these loans, creating the potential for loan originators to sell “lemons” (in the Akerlof sense) to downstream loan buyers while retaining the best loans in their own portfolios. Poppo and Zenger (1998) recently provided evidence on such measurement costs as they affect both the choice of vertical scope and the resulting satisfaction with the governance choice.

While uncertainty about product quality gives rise to the lemons problem, there are additional transaction risks created by uncertainties in the business environment. In particular, firms may be reluctant to make long-term relationship-specific investments when there is the possibility that these investments can be expropriated ex-post (Williamson 1975, 1985; Klein, Crawford and Alchian, 1978), a problem that is often not easily remediated by contracting (Hart, 1995) when the environment is uncertain. Given that efficient production and trade often requires investment in assets unique to a relationship by the buyer, the supplier, or both parties, the possibility of ex-post opportunism decreases efficiency of trade (by discouraging investment) and can eliminate trade opportunities altogether. This perspective has been extensively researched and documented (Shelanski and Klein, 1995; Williamson, 1999; Boerner and Macher 2003), with empirical studies generally supporting the notion that asset specificity and uncertainty correlate with integration.

In the mortgage banking sector, the principal productive assets (human capital, office space and information technology) are largely fungible across trading partners due to the standardization of
products and transactions in the industry and the large number of potential counterparties for trade. Whereas in the early 1980’s there was some co-specialization along the value chain, the emergence of credit scoring, automated underwriting engines, and other types of information technology has reduced these risks (see Jacobides, 2005). In other words, the most significant TCE-oriented risks are ex-ante problems of information asymmetry, rather than ex-post risks of opportunism. Thus, in our setting, TCE theory leads us to posit that:

Hypothesis 1: The greater the information asymmetry and the greater the problems of assessing the quality of what is traded, the greater the extent of vertical integration.

The Proposed “Comparative Advantage” View

An alternative line of reasoning, which has emerged more recently, has to do with firms’ particular resources and capabilities. This sub-section summarizes the relevant background and related literature and then briefly describes our comparative advantage perspective on integration (see Jacobides and Winter, 2005, for a more detailed exposition).

Existing Research linking Capabilities and Vertical Scope. Over the last few years, it has become clear that TCE is not a self-sufficient theory of vertical scope. Williamson, for instance, recommends that the traditional TCE query “‘What is the best generic mode (market, hybrid, firm) to organize X?’ be replaced by the question ‘How should firm A -- which has pre-existing strengths and weaknesses (core competences and disabilities) -- organize X?’” (1999: 1003). This question has been recently pursued by Madhok (2002), who suggested that an individual firm’s choice must depend not only on the characteristics of the transactional conditions, but also on its strategic objectives, the attributes of its own capabilities, and the governance context it has created. On this broader view of firm scope, there has been significantly less empirical evidence.

While Walker and Weber (1984) found that the most important predictors of sourcing were cost differences between the focal firm and outside suppliers in producing a specific component, until recently, most of the subsequent research in firm boundaries did not include firm-specific cost or
capability measures, or use capability differences as controls. More recently, progress has been made. Argyres (1996) was one of the first to provide qualitative evidence on the role of firm capabilities in integration decisions, observing that in the cable manufacturing business capabilities were a significant driver of vertical scope, in addition to transaction costs. The growth of the capability- and resource-based view of the firm further led scholars to consider the role of firm heterogeneity. Skill sets (Poppo and Zenger, 1998), specific experience (Leiblein and Miller, 2003) and capabilities (Hoetker, 2005; Jacobides, 2005) have been show to be important determinants of outsourcing decisions. Path-dependency in boundary choices has also been attributed to firm capabilities in contracting (Argyres and Liebeskind, 1999; Madhok, 2002).

The “Comparative Advantage” Theory of Vertical Scope. To understand the role of capabilities in the choice of governance structure it is important to note that the “market” is really is an organizational interface, behind which is another firm -- a firm which finds it advantageous to sell on the basis of its own productive capabilities. Classic TCE tends to treat the market as a collection of undifferentiated firms, a condition that would be true of a competitive market with identical technology (Nelson, 1991). Yet once we allow for differences in the productive abilities of different firms, the relevant choice is between the firm, and its cost levels (if the input is produced) as opposed to the cost levels of another firm, including its profit margin, after deducting transaction costs (Jacobides and Winter, 2005).

To make headway in a theory of capability-based analysis of vertical scope, we also must further refine the concept of capability. An important distinction we introduce in this paper is between productive capabilities of a firm (productive efficiency or “zero-order capabilities” – see Winter, 2003) in each of the vertically related stages, and the capabilities of governance, the ability of a particular firm to use integration or the market to create value by linking these two stages. To be precise about this distinction, we introduce some notation. Consider a firm that performs an upstream activity (e.g., production) and is considering whether to insource or outsource a downstream activity (e.g., sales). Let the efficiency of the downstream activity for the
focal firm \((i)\) be represented by \(p_i\), and the bureaucratic cost of integrating both activities internally, including the cost of muted incentives, as \(B_i\). Similarly, let the productivity of a potential outside vendor \((j)\) be represented by \(p_j\) and the transactions cost for firm \(i\) procuring product from that vendor \(j\) through the market be represented by \(TC_{ij}\). In this framework, the vertical integration decision depends on the comparison of whether: \(p_i - B_i > p_j - TC_{ij}\).

Historically, the TCE literature has excluded productivity differences of the focal firm and the outside firm (vendor), that is, \(p_j - p_i\), from its direct purview, and has considered internal governance costs \(B_i\) as a constant, non-negative term (Williamson, 1985: Ch 6). Instead, it has focused on the drivers of variation in transaction costs. The more recent TCE literature has, in addition, recognized the impact of firm-specific factors by considering additional variation across firms in \(TC_{ij}\). Such research has argued that transaction costs vary either across suppliers \((j)\) or across buyers \((i)\). For instance, Dyer (1996) argues that using particular suppliers consistently across time mitigates the risks of procurement from that set of suppliers \((j)\). A number of studies (Winter, 1988; Argyres and Liebeskind, 1999; Leiblein and Miller, 2003; Mayer and Argyres, 2004) argue for a path-dependent development of governance capabilities at the buyer \((i)\) leading some to be better at market procurement. Such studies, though, have typically not focused on productivity differences—they address variation in the capabilities of governance, instead of variations in productive capabilities (Demsetz, 1988; Langlois and Foss, 1999).

A focus on productive capabilities as drivers of vertical scope, though, only becomes meaningful if there is reason to believe that there is variation in productive capabilities across firms. While the limited evidence on intra-industry productive capabilities or cost structures seems to corroborate this view (Lieberman and Dhawan, 2001) there are good theoretical grounds to expect that this should be so. Capabilities are heavily driven by the firms’ general and specific knowledge of the production process (Richardson, 1972; Teece, Pisano and Shuen, 1997). This
knowledge is developed by a path-dependent process of complementary investment and learning-by-doing (Barney, 1994, Winter, 1995), shaped by the numerous contingencies firms face in their operations (Winter, 1990, Levinthal, 1997). To the extent that process knowledge is sufficiently complex to limit imitation (Porter 1996, Rivkin 2001, Siggelkow 2001) or include irreversible investments (Ghemawat, 1991; Winter, 1995) that alter future opportunities, these differences can be sustained for substantial periods of time. Moreover, in a volatile environment, differences in capabilities can be further accentuated by differences in the ability of organizations to learn and adapt (Teece, Pisano and Schuen, 1997). Collectively, these arguments suggest that even in an environment where primary resources are homogeneous, different organizations are likely to display significant variations in processes leading to differences in productive capability.

The more pertinent question for our context is how the productive capabilities of a firm are distributed along the value chain, which in our analysis directly influences the integration decision. In an industry with two vertical segments (upstream and downstream) a firm which has strong capabilities upstream and downstream, will be integrated. A firm that is stronger in one segment (say, upstream) may find it advantageous to specialize in this activity and utilize the market for the other activity at which they are disadvantaged (e.g., sell to a downstream firm). Thus, specialization by using the market will happen to the extent that some firms are good upstream and some downstream.

This observation leads to two specific predictions about vertical scope. First, the more capable a firm is in a particular segment, the more likely that the firm will participate in that segment. This arises in part because there are limited gains from trade to contracting for this activity, and partly because a firm may be able to expand in activities for which they are advantaged. This suggests the following hypothesis:

Hypothesis 2: Superior productive capability in a particular vertical segment will be positively associated with activity in that same vertical segment.
Hypothesis 2 argues that the better a firm is in a segment—say, downstream, the less it will tend to use the market as opposed to its own downstream operations, as doing so would be relying in inferior capabilities for that segment (since $p^D_i > p^D_j$). An upstream firm is more likely to be integrated into the downstream segment, if it has a superior downstream productive capability.

We further argue that superior capabilities in one segment will be associated with a smaller degree of integration into the other vertically linked segment. The better an upstream firm is in the upstream segment, the smaller the probability that it is vertically integrated into the downstream segment, all else being equal. That is, the better a firm is upstream (i.e., $p^U_i > p^U_j$) the greater its use of the intermediate market linking the up- and down-stream segment, ceteris paribus: Firms capable upstream will want to use the intermediate market as a net seller of upstream goods, and as such are more likely to be less vertically integrated. This means that a comparative advantage in one segment leads, other things being equal, to greater tendency to be vertically specialized. The objective is to leverage the productive capability $p^U_i > p^U_j$. Thus we expect that:

**Hypothesis 3:** Superior productive capability in a particular vertical segment will be negatively associated with integration in its downstream segment.

Our theoretical perspective thus provides a consistent explanation for why firms may be active in both vertical segments, as opposed to simply specializing in the segment where they have an advantage. This may explain why we regularly observe “mixed governance choices”, an anomaly for much of the existing theory (see Harrigan, 1985; 1986; Parmigiani, 2004).

To illustrate how our approach explains mixed governance, consider a firm with a better-than-average upstream division, and a worse-than-average downstream division. The fact that it has a downstream division at all may be due to historical reasons, perhaps reflecting different competitive conditions, or even past mistakes. Yet it may have an economic incentive not to redress them, by abandoning the relatively inefficient downstream division: The irreversibility of prior commitment downstream (Ghemawat, 1991) or the fact that the costs for operating
downstream are sunk makes it economical, on the margin, to maintain the original downstream operation as a going concern. We also expect that such a firm will grow its upstream operations, and shrink its downstream segment. Note that the quest for superior rates of return (as opposed to absolute profits) may encourage a firm that is above average in both segments but better in one to grow more in its area of comparative advantage. Thus, the existence of sunk investments and limits to the speed of entry or exit in a segment underpin Hypotheses H2 and H3, and can explain both “mixed governance” and the gradual changes in firms' scope often observed in practice.  

Our analysis thus relies on comparative rather than absolute (‘best in class’) advantage. For specialization to occur, the gains from trade must simply outweigh the cost of using the market (the “net tax” of TC). Indeed, we predict that all firms for which gains from specialization surpass transaction costs are likely to be specialized whether or not they are the most capable firm in a particular segment. This further underscores the importance of heterogeneity in capabilities in the industry -- gains from trade can only happen if productive capabilities are not symmetrically distributed along an industry’s value chain.

The Comparative Advantage vs. the Resource-Based View. Our perspective is related to but distinct from the Resource-Based View (RBV) of vertical scope. In the RBV, integration can be a natural consequence of an attempt to exploit a unique and non-tradable asset in the pursuit of competitive advantage (Combs and Ketchen, 1999; Afuah, 2001; Schilling and Steensma, 2002; Leiblein and Miller, 2003; Leiblein, Reuer and Dalsace, 2003). The variation in the ownership of these assets leads to variety in integration, depending on the relative costs acquiring and integrating the requisite complementary resources and capabilities (Barney, 1999). While capability differences do play a role in the RBV (White, 2000; Leiblein and Miller, 2003), the emphasis of the theory is on how vertical scope changes to exploit particular resources. Our analysis takes a “simpler” view, that any capability differences in the value chain, even a potentially replicable or short-live comparative advantage, may explain a significant part of the variance in the way firms choose their vertical scope.

10
That being said, the theory we develop in this paper focuses on a fairly narrow, static picture. We do not explicitly consider the cost of developing capabilities or at whether the capabilities are best formed organically or through mergers and acquisitions; all we argue is that at any point in time, firms decide their scope on the basis of their relative, *comparative* and not necessarily absolute advantage. We return to the limitations of our argument, and to the necessary extensions of this research approach in the discussion.

To recap, different theories have shed light to what drives vertical scope. Table 1 summarizes the major approaches, identifying their focus and predictions as for when we expect to see integration, as a function of the attributes of the transaction (information asymmetry; threat for opportunism in a re-negotiation under asset-specificity), or of the transactors (concentration in their markets; capabilities in carrying out transactions; etc.).

METHODS AND DATA

Research Design and the Choice of our Setting

*Research Design.* Our general analysis approach is to consider the relationship between integration and the factors driving integration as predicted by the measurement branch of transaction cost economics (H1) and our “comparative advantage” perspective (H2, H3). We are not only interested in testing these hypotheses, but also examining their relative ability to explain the variance in integration across firms.

*Background: Mortgage Banking.* Mortgage banks are non-depository financial institutions that originate, process, approve, and then (in most cases) sell mortgage loans to the secondary market through securitizers such as Fannie Mae. The mortgage banking industry is a good setting to study vertical scope, as it has a complex value chain which has increasingly been fragmented into quasi-independent parts, which are performed both by integrated firms as well as narrowly targeted specialists. The structure of the mortgage banking industry is provided in Figure 1A, which shows...
the vertically co-specialized world of mortgage banking (which itself competes with more integrated Savings & Loan institutions of banks that also offer mortgages). The industry has a nearly 1300 firms over our sample period, active in different combinations of value chain segments, which provides both a large and diverse sample, and also eliminates issues related to small numbers bargaining (which would require additional TCE-related measures). Mortgage banks account for more than 56% of the total mortgage loan production in 2002, about $800 billion in new loans. Mortgage banking is an important, even if under-researched sector.

Mortgage banking is one of the least regulated segments of the financial services sector, perhaps because the mortgage origination process carries little systemic risk for the financial services sector as a whole. As a result, mortgage banks are free to make their own choices of scope unaffected by regulatory pressures, and thus our dependent variable, the degree of integration, represents a true firm choice. However, while there is no formal regulatory framework for the industry, mortgage banks are called upon by their Association and the securitizers to disclose key measures of financial and operational health through a survey. This database offers unusually detailed measures in different parts of relative efficiency in each part of the value chain.

Our particular focus is on the most upstream part of the mortgage loan production process, which is illustrated in Figure 1B. Specifically, we look at one of the most important decisions of vertical scope for mortgage banks: the question of whether they are integrated in the production of loans, or whether they use correspondent banks or loan brokers from whom they procure loans. Mortgage banks can outsource the generation of the lead, the gathering of the relevant paperwork and the filling of the relevant forms, as well as the hand-holding of the loan applicant to a mortgage broker or to another bank (a “correspondent bank”). The broker or correspondent would then “sell” the qualified loan to the bank (receiving a commission ranging from 80 to 150 basis points depending on how close the loan is to funding, and how valuable it is). This choice, then, is a clear make-vs.-buy: The bank can make the loan itself, or buy it from another entity.
Figure 1B also provides an illustration of what exactly we mean by the term “productive capabilities” in our setting. We focus on two vertically linked segments: Loan origination, and loan warehousing. Productive capabilities in loan origination determine the efficiency with which a mortgage bank or a mortgage broker undertakes the steps described in the steps described in Figure 1B: Pre-qualification, document generation, application processing, credit analysis and underwriting, and approval. Productive capabilities in warehousing affect the performance in engaging in pipeline / warehousing analysis and the transfer to the secondary market (ability to manage interest rate risk, prepare the loan bundles for sale to secondary market investors, etc). Governance capabilities affect the ability of the firm to link the loan origination process to the loan warehousing process. Each of these segments has well defined activities, volume measures and productivity metrics enabling them to be evaluated independently.

**Data: Mortgage Banks’ Integration into Retail Loan Production**

*Retail Integration as a Dependent Variable.* We take the perspective of all banks that warehouse a loan and consider whether these banks make their own loans, whether they buy them, or do both. Mortgage banks can either be fully integrated (never using outside agents); fully specialized (a type of bank that became known as a wholesaler); or use a mixed mode, that is both buy and make loans. Most banks in our sample are indeed users of the mixed mode and thus the degree of integration varies continuously from 0% to 100% specialized in origination. This continuous, measurable variation in integration provides a significant measurement advantage as it avoids statistical inefficiency from forcing a naturally continuous choice (cf. Harrigan, 1985) into a binary (“make” vs. “buy”) measurement approach as is common in TCE studies.

*Data: The Mortgage Banking Financial Reporting Form.* Through collaboration with the Mortgage Bankers Association of America (MBAA) we gained access to the confidential and quasi-regulatory database that the MBAA maintains on the activities of the mortgage banking sector, called Mortgage Bankers Financial Reporting Form (MBFRF). Each year the MBAA
selects a sample of mortgage banks to survey with the goal of accurately capturing the diversity of activity in the industry, avoiding potential self-selection biases, and covering a reasonable share of the total market. The sample is an unbalanced panel of all firms that have warehousing activities, and the yearly observations range from 79 to 285 per year. Each firm is sampled for an average of 3.1 years, out of a possible maximum of 10 years, and the total number of firms in the sample is 685. On average, the respondents of the MBFRF are responsible for about 25% of the total loans produced in the United States, and as such this is a highly representative sample. The database we utilize covers the period 1988-1998, and contains 1,792 usable observations.

The MBFRF contains detailed data on the costs, revenue, output quantity, and employment in each part of the production process enabling direct measures of productive capabilities by value chain segment, degree of integration, and firm scale. In addition, characteristics of the loans produced or managed are also captured, which enable the measurement of transaction costs.

**Measures: Integration, Transaction Costs, Comparative Advantage Metrics and Controls**

*Dependent Variable: Integration in Retail.* Our dependent variable is integration in retail loan production; more specifically, the percentage of total loans produced through a bank’s own retail branches (as opposed to those purchased from brokers or correspondents) in terms of dollar value.

*Transaction Costs.* From our industry interviews we determined that the principal risks of using the market for the procurement of loans surround the risks of trading loans of unknown quality. Borrowers can vary in ways that are difficult to observe or objectively measure that influence their probability of delinquency or default – a very real cost to the mortgage banks.² Originators have a strong incentive to “disguise” the true riskiness of the loan by the way data are reported and to sell loans with the same observable characteristics that they believe have a higher default risk. The market cannot easily correct for this behavior due to the significant time lag between origination and default. We conjecture and validate empirically that the riskiest loans categories in terms of likelihood of default also have the most unobserved risk. Government-sponsored loans (FHA/VA)
are known to be the highest-risk type of loan due to the large heterogeneity of the eligible pool of borrowers, and the fact that the credit standards are more generous as a matter of social policy. “Plain vanilla” fixed-rate conventional mortgages (FRMs) have intermediate levels of risk -- their long-term and (relatively) high interest rate raises the possibility of default due to unexpected events. Finally, adjustable rate mortgages (ARMs) have the shortest expected term and lower interest rates, making them the least risky, as industry-wide default rates suggest.

To confirm that this categorization was indeed indicative of transaction risks we conducted several tests (see the Appendix). Overall, we find that firms that are integrated face lower default risk overall, that the default risk for different loan categories follows our discussion of relative risk of various loan categories (FHA/VA loans are riskier than fixed-rate loans), and that high risk loans procured through the market indeed have higher default rate. These observations broadly validate our use of loan categories as a measure of risk and show that these risks relate to integration. Our measure of transaction costs, then, is the composition of the loan portfolio in terms of percentages of these three loan types – firms with a larger proportion of “dangerous” loans should be more integrated in order to mitigate such risks (Hypothesis 1). Measuring TC in this manner has the advantage that it represents a direct measure of transaction risks relevant this setting; the shortcoming is that because “hold-up” is not a major risk in this sector, we cannot make any generalizations about the relative explanatory power for these other TC risks versus our hypothesized productive capability factors.

Measures of Productive Capability. We measure capabilities by labor productivity (output per labor input) and operating margin (revenue less total cost) consistent with academic research in productivity and industry practice (Posner and Nambiar, 2002). Since outputs are relatively standardized in this industry and the MBFRF takes great care to properly capture revenue, staffing and total cost for each segment, we believe we have accurate measures for these constructs for both the origination and warehousing segments. We define labor productivity as the number of loans processed (originated in origination, serviced in warehousing) originated per full time
equivalent employee (FTE) in each segment. Results are similar whether we use loan count or dollar volume. Operating margins obtained in each segment through the MBFRF income breakdown which isolates segment specific cost and revenue. While measures of per-segment margin can be noisy, especially in financial institutions, they do provide useful information that can be used to corroborate the FTE-based efficiency metrics. Given the distributional properties of the margin and of the FTE per loan metrics, we used the natural logarithms of these measures in our regression. Our hypotheses suggest that integration into retail should be positively related to efficiency in origination (H2) and negatively related to efficiency in warehousing (H3).

*Control Variables:* We include control variables for time (a year dummy variable) to control for changes in the pool of respondents year to year, which could alter the sample means of all variables, and to eliminate business cycle effects. For instance, in years with high demand, mortgage banks are known to use more outside agents, since their own branches are not enough to cover the necessary demand (Nishiguchi, 1994). We also control for scale to avoid confounding our analysis of integration with simple economies of scale or market power arguments which predict that size in a segment and efficiency should be related. Our primary measures of scale are assets and assets squared, although results are unchanged when we measure size as loan volume or employment.

*Methods*

We report four types of results. First, we report ordinary least squares (OLS) with all firms and years pooled, estimated with Huber-White robust standard errors accounting for repeated observations of the same firm. Second, we corroborate this analysis using robust regressions (Cook, 1977; Berk, 1990; Hamilton, 1991) which performs an iterated procedure that reduces the overall influence of outliers, a common problem in financial industry data since assets and returns can vary more than in “industrial” corporations. This is particularly useful as our confidentiality agreement prevents us from identifying the actual firms which would enable a case by case
investigation of the outliers. Third, we employ a panel data fixed effects model (cf. Baltagi, 1995) to address firm heterogeneity, which effectively controls for any time-invariant firm specific factors in our unbalanced panel. However, our confidence in these models is guarded, as during the observation period there has been substantial merger and acquisition activity, and as a result fixed effects are indicative rather than conclusive. Finally, we use logistic (Logit) regressions (again with Huber-White standard errors) to compare our analysis to the binary choice (integrated or specialized) more common in the TCE empirical literature. For this final analysis we limit our analysis to firms that are nearly integrated or nearly specialized (greater than 90% or less than 10% activity in origination). Qualitatively similar results are obtained using other cutoff points including 0% and 100%.

In all four methods, we sequentially introduce explanatory variables for scale, for transaction cost related factors and finally capability related factors. The consistent results of multiple methods (as well as more specifications we tried in addition to those reported) indicate significant robustness in our analysis. Finally, as we note in the research design section, we are not only interested in the statistical significance of the coefficients associated with each theory; we are also interested in the extent to which the measures can explain the variation in our dependent variable.

RESULTS

Model 1 in Table 3 presents the correlation between the degree of integration, control variables, and scale. We can see that our control measures do not drive the results: Rather than the predicted positive linear relationship between scale and integration, there is a “u-shaped” relationship, with a strong and undoubtedly negative linear term, and a positive squared term. Integration is neither linear, nor monotonically positive function of scale. Similar results obtain in the Logit regressions of Table 4, although they are less strongly significant.

Include Tables 2, 3 and 4 about here
Hypothesis 1 (on the role of TC) received mixed support, as we can see through Model 2 that includes controls and the TC related measures. In the logistic regression, which considers what explains the choice of firms to be entirely integrated or entirely specialized, the sign and magnitude of the coefficients are what the theory would predict (see Table 4). Relatively safe loans (FRM) are related negatively to the decision to be fully integrated, and intermediate-risk loans such as ARM are not statistically correlated to integration. Finally, we find that the greater the proportion of risky loans a firm produces (FHA/VA), the greater the probability that the firm would be integrated, although this finding becomes significant only in the full model.

The picture changes once we consider not only the binary choice of integrating versus not integrating, but also the precise level of integration for the large number of firms that rely on both the market and their own retail branches. Model 2 in Table 3 contains the results for the OLS, robust and fixed-effects regression on the continuous measures of integration. In OLS, the composition of the loan portfolio is not statistically significant in explaining integration. However, in our most restrictive specifications, robust and fixed effects regressions yield results that occasionally contradict the transactional logic.6

What is more interesting to us is the amount of variance explained. Even in the Logit regression, where the results go in the expected direction, the increase in fit (pseudo-$R^2$) is 2.6% (from 1.6% to 4.2%). To put this in context, the addition of capability metrics increases the pseudo-$R^2$ to 72.6% (see below). The ability of transactional factors to explain integration is even weaker in the continuous measure of integration, where the fit increases from 5.1% to 5.7% in OLS, and the within-$R^2$ in fixed effects increases from 4.2% to 5.7%.

Hypotheses 2 and 3 (relating to productive capability) are very strongly supported as we can see by looking at the results of Model 3, which includes the productive capability measures. Table 3 provides the results for the continuous measure of integration. As we can see, the capability-related coefficients in terms of productivity (FTE per loan) are significant in the expected
direction. Thus, the greater the efficiency downstream (the smaller FTE in origination), the greater the extent of integration into that segment. Firms that are better than others in origination are integrated in that vertical segment (H2). Furthermore, the coefficient of the efficiency in the upstream part of production (loan warehousing) is strongly negatively associated with integration (H3). Similar findings obtain for the margin measures we use. The higher the origination margin in a segment, the more that segment is used (as opposed to the market) as predicted by Hypothesis 2. Also, higher warehousing margins are associated with a greater the reliance on the market (H3). Again, this dramatically increases explanatory power; in OLS, the $R^2$ went from 5.6% to 50.5%.

These findings are corroborated in the robust regressions. Indeed, margin in retail production, the only capability metric that was not above the significance threshold, becomes clearly significant. Also, scale effects almost vanish in the robust specification, once we include the capability metrics. This further suggests that our results are not confounded by scale issues.

The fixed-effects analysis is consistent with the OLS and robust regression in supporting H2 and H3 --most of the coefficients are highly significant in the same direction. The reduction of significance was expected in a fixed-effects specification since it removes any fixed long-run components of firm capability, leaving only the transient component. The results of this regression give further credence to the arguments supporting H2 and H3 as they suggest that capability differences and integration not only covary between firms, but also over time for the same firms. However we are somewhat guarded about this analysis since the fixed effects estimates can be contaminated by changes in the nature of a firm resulting from M&A which is common on our sample.

The Logit analysis is also consistent with OLS, Robust, and Fixed Effects. The coefficients for both productivity and margin are significant and in the hypothesized direction, except for the warehousing margin measure (H3b). In addition, when we add the capability-based metrics, pseudo-$R^2$ in the Logit regression increases from 5.6% to 72.6%.
DISCUSSION

This analysis suggests that, at least in our particular setting, differences in productive capabilities are key drivers of the decision to integrate. While a transactional logic may be correct (at least in determining the binary institutional choice), capability considerations dominate. While this finding may relate to the specificities of our setting, they do not undermine the major premises of our study, that 1) productive capabilities can and often do play a major role in the determination of vertical scope, and also account for “mixed governance mode”; and 2) that comparing the relative explanatory power of different perspectives on the determinants of vertical scope can yield new insights. This suggests, at a minimum, that there may be significant value in utilizing measures capturing non-TCE factors that determine scope in future empirical work on outsourcing and firm governance decisions.

Our findings lend direct support to the capabilities view of vertical scope (Jacobides and Winter, 2005). Specifically, our results are consistent with prior studies that showed that the make-versus buy choice is affected by firm specific factors that can be interpreted as correlates of firm capabilities (Walker and Weber, 1984; Leiblein and Miller, 2003; Hoetker, 2005). We also distinguish between the productive capabilities, which have been scarcely studied, and the capabilities of governance, which have recently received more attention (Dyer, 1996; Madhok, 2002; Mayer and Argyres, 2004). Our study finds that productive capabilities are perhaps the most important driver of vertical scope. We also find that comparative advantage upstream may drive the concurrent use of in-house production and market arrangements downstream, and that the mix of in-house and market-based production downstream depends on the balance of capabilities.

These findings open up a new line of inquiry. As we know from the surprisingly limited empirical evidence (e.g., Lieberman and Dhawan, 2001) variation in capabilities is persistent and important within an industry, and, in all likelihood, within firms in the value chain (cf. Duncan, 1998). As such, we should expect that many settings would be much like ours, where comparative advantage happens to play a major role in predicting vertical scope.
Our findings, then, show that in examining vertical scope, we have to consider what is “behind” the market; and that, over and beyond the nature of the “interface” used to govern a transaction (in-house coordination or market procurement), there is some organization or entity “doing the work”, i.e. turning inputs into outputs at a stage of the production process. Simple logic suggests, yet much research neglects, that if a firm has a comparative advantage in a particular part of the value chain, another firm has a corresponding weakness. For gains from trade to exist, there must exist a symmetrical capability imbalance, so that both parties gain from specialization. This focus on what drives comparative advantage is also subtly different from recent applications of the RBV to vertical scope decisions (Combs and Ketchen, 1999; Schilling and Steensma, 2002), which consider the role “absolute” advantage. We argue that the patterns of specialization are largely driven by comparative advantage; and that to understand comparative advantage we have to understand the distribution of capabilities along the value chain.

This proposed focus on the systematic patterns of productive capability distribution opens up exciting new venues for research. The questions thus become, what gives rise to potential gains from trade and specialization in an industry? Is heterogeneity of capabilities and the resulting latent gains from trade a function of different knowledge bases (Winter, 1987) along the value chain? Or are such latent gains just the result of historical happenstance, re-inforced by path dependence and selection? Or are there, in particular settings, real managerial diseconomies of scope, so that being in one part of the value chain adversely affects performance in another? (cf. Jacobides, 2005) That is, might vertical specialization be related to inappropriate use of common HR and organizational practices or to the establishment of firm-wide “rules of thumb,” or simple rules, common across the firm (Eisenhardt and Sull, 2001)? Is it due to the establishment of “dominant logics” (Prahalad and Bettis, 1986), simplifying devices (Gavetti, Levinthal and Rivkin, 2003) and mental models (Markides, 2000), which help one part of the value chain and hinder another, leading to the intra-organizational benefits of differentiation (Gulati, Lawrence and Puranam, 2005)? Asking such questions and understanding the distribution of comparative
advantage will lead us to re-integrate the theory of production and the theory of exchange (Demsetz, 1988; Langlois and Foss, 1999), and build a more robust set of explanations of the institutional structure of production (Coase, 1992; Madhok, 2002; Jacobides and Winter, 2005). The central question becomes: When and why do productive capabilities not co-vary?

Yet while we consider that differences in productive capabilities are very significant and should be studied further, we also believe that this analysis should not be done to the exclusion of the analysis of transactional factors. On the empirical level, we find that transactional factors are important. More importantly, on the theoretical level, we consider that transaction costs and capability differences not only work as independent motivators of vertical scope, but, crucially, they interact to shape the Institutional Structure of Production (Coase, 1992). TC determine the extent to which these latent gains from trade, driven by the distribution of productive capabilities in an industry, are or are not sufficient to justify vertical specialization. Much as global taxation can curb international trade, costs of using the market may prevent specialization and reliance on the market. So TC are such that they catalyze the differences in capabilities; their reduction enables the underlying differences in productive capabilities to be brought to the fore, and as a result, allows specialization to occur. However, it is not the reduction of transaction costs itself that matters but the relative reduction in these costs, when compared to the capability differences that lead to the latent gains from trade.

Finally, another connection between productive capabilities and TC is that the very reduction of TC may be motivated by the existence of latent gains from trade. Interested parties, which stand to gain from specialization, may push and invest for TC to be reduced, expecting their investments in reducing TC to be repaid through their share of the gains from trade (Jacobides, 2005). The joint analysis of differences in productive capabilities and transaction costs / vertical scope holds much promise for the future and opens up new venues of research.
Limitations

While this study does indicate that productive capabilities matter, it is subject to a number of limitations. On the empirical level, our study does not directly test Williamsonian TC. While we do believe that in our setting the appropriate TC are those related to misrepresentation and risk and that these are well captured by our measures (both theoretically and through empirical validation), we still do not directly test for the relative ability of Williamsonian TC to explain variance. So our findings cannot rule out that further unobserved TC are empirically as important as capability differences in explaining scope, especially in other settings. That being said, the measures we used for capability differences cannot be reasonably construed as indirectly associated with Williamsonian TC, so that our basic finding, that is, that productive capabilities matter, and explain a large part of the variance, still stands.

On the theoretical level, there are additional limitations. First, our paper presumes the existence of a clear, separable governance interface; that is, it is built on the assumption that there is a potential market, with some level of TC “tax”. While this type of segmentation appears in the industry we consider, it may not broadly apply to industries where boundaries between segments vary. Indeed, a key analytical challenge which we did not tackle is to examine why some segments do not even have the option of being vertically distinct. While we would argue that there is something in the nature of capabilities that is associated with what is often called “technological inseparabilities”, our approach both takes them for granted, and cannot explain them. We consider recent research on this topic (Baldwin and Clark, 2003; Jacobides, 2005) to be an important complement to our approach. Similarly, the emphasis of our theory is on the relationship of capability differences and scope rather than how these capabilities differences arise. There is much promise in examining how productive capabilities and governance capabilities affect and determine each other over time (Argyres and Zenger, 2004); and how capabilities and transaction costs co-evolve, shaping the structure of the production system (Jacobides and Winter, 2005).
Furthermore, while recent research (e.g., Gulati, Lawrence and Puranam, 2005; Jacobides, 2005) suggests a theoretical basis for capabilities to vary across firms, it does not necessarily explain why capabilities may be vastly different in some settings while very similar in others. Thus, while we do not offer a comprehensive theory of vertical scope, we hope that pointing out an important empirical regularity (that capability differences, at any point in time, drive scope), as well as articulating the comparative advantage view (albeit in a static sense) will help support future research in scope, capabilities, and their co-evolution.

Second, while our objective was to provide a more inclusive theory of vertical scope, we also did not directly consider any of the “dynamic” drivers of integration. In the theory section, we noted that a motivation for vertical integration is the protection or access to a stream of revenues from upstream innovation. This might not be particularly relevant in a mature industry like mortgage banking, yet it would be a significant omission in a more dynamic setting (White, 2000; Afuah, 2001; Schilling and Steensma, 2002; Leiblein and Miller, 2003).

Finally, our paper takes the vantage point of the entire industry, looking at the logic that drives specialization; it tries to shift from the “ceteris paribus”, partial equilibrium analysis inherent in TCE as well as the RBV, to a systemic analysis of what drives specialization in a given population of firms with a particular distribution of transaction costs and capabilities. However, while such an approach yields some fresh theoretical insights, it does not enable us to speak directly to the problem a manager faces in determining the scope of their firm. In particular, our approach is silent on the cost of developing capabilities; and it does not examine whether either organic growth or M&A can change the capability distribution over time and how a firm might balance these strategies to close capability gaps (cf. Karim and Mitchell, 2000). We consider these topics as important extensions of our research. On the basis of the knowledge that capability differences do affect vertical scope, we hope that future research will examine why in some settings firms manage to balance their capability gaps, precluding specialization; and whether this happens organically or not. Such an approach, building on our findings, could potentially expand on the
important recent research by Helfat and Raubitschek (2000) and Karim and Mitchell (2000), applying these author’s insights to the question of vertical scope as it evolves.

**Concluding Remarks**

This study proposes that in the determination of vertical scope, we have to consider separately the capabilities of governance of different governance interfaces (markets vs. integration) as well as the productive capabilities of the firms in an industry. It argues that behind “the market” lies another firm, willing to supply a good or service at a particular price; and that the ability of any such firm to produce in a cost-effective manner will be an important part of the calculus of the make-vs.-buy decision. In a world of significant variation in these productive capabilities between firms, capability differences may be responsible for a significant part of the decisions to integrate or not -- exactly as it happened in our setting. Transaction costs can thus be seen as a tax that seriously impacts, but does not in and of itself generate, the need to specialize.

Our deceptively intuitive finding, i.e., that the choice of scope is largely determined by comparative capabilities, opens up some fascinating questions, once we shift the focus from the individual firm to the population of firms and the related distribution of capabilities. The analysis of the co-variance of productive capabilities of different firms along the value chain should itself become the object of study, which can help us understand the deeper causes of vertical scope. By expanding TC analysis and combining it with a more careful study of the managerial structure of production, we will be able to obtain a more robust and more representative theory of vertical scope. Building such a theory, and focusing on the ability to explain variance, as opposed to the correctness of a particular theoretical prediction can help us not lose sight of the forest for the trees.
References


Endnotes

1 A related rationale for mixed governance rests on “differential scalability” between vertical segments – that is, on the possibility that it is costlier to expand, say, the downstream operations than to expand the upstream operations. Consider a firm with two equally strong vertical segments, which wants to grow and leverage its absolute competitive advantage: As it can grow upstream faster and more economically than it can grow downstream, it will eventually both produce downstream, and use the market, so as to benefit from its superior and more scalable upstream capability (Jacobides, 2004).

2 Loan defaults are costly for mortgage banks in several ways. First, if a loan defaults and there were underwriting errors, mortgage banks are often obligated to buy back the loan from secondary investors and bear the credit loss directly. Secondly, even if the mortgage bank is not responsible for credit risk (for example, due to a government guarantee), managing default creates significant operational costs of collections, foreclosure and subsequent asset management. Thus, the costs of procuring a “lemon” from the market are large.

3 We utilize labor productivity rather than multifactor productivity which is also commonly employed in performance and productivity analysis. In this setting, the two measures are likely to behave similarly as production capital (computers and office space) is generally proportional to labor. Thus, any conclusions we draw from a labor productivity analysis are likely to apply to a multifactor productivity analysis as well. We are unable to empirically verify this as the MBFRF does not include an inventory of non-financial capital.

4 Given that we measure integration as a proportion rather than absolute size, this may be less of an issue since scale arguments generally relate to overall size.

5 We utilize the rreg procedure in Stata 8.1 to perform this analysis. Results are similar when we utilize median regression.

6 The reader might observe that the coefficient signs on the composition portfolio of the loan portfolio do change between different models and between the logit and OLS / robust / fixed effects regression. This, however, is an expected statistical artifact since all loan categories add up to 1: A negative coefficient in, say, FRM and a positive coefficient in FHA/VA is equivalent to having a positive coefficient in both FRM and FHA/VA inasmuch as the beta for the FRM coefficient is smaller than the one in the FHA/VA. In both of these cases a shift from FHA/VA to FRM will lead to reduction in the level of the dependent variable (i.e., an increased reliance on the market).

7 If unobserved heterogeneity in transactional conditions (relating to differences in the leeway for ex post, non-contractible bargaining) drives scope, this “additional” explanation will help reduce the unexplained residual. Yet the variance explained by capabilities is fairly substantial. Thus, the ability of contractual / hold-up conditions (unrelated to measurement problems) to explain variance is limited in our setting.

8 In our setting (and, we suspect, many others) it is hard to construe a scenario whereby a strength of a particular firm in one segment (e.g. its efficiency in warehousing loans) is a surrogate or a correlate of its ability to guard itself from renegotiation with another firm downstream. The only potential causal link that might exist is that the strength of a firm in a particular segment (e.g., a superior warehousing capability) reflects the “savings” associated with adopting a superior but transactionally hazardous technology. This does not seem to have any face plausibility in our setting, yet we did explore it econometrically. Specifically, we examined if integrated firms (which by definition do not face transactional risks) are more efficient upstream than non-specialized ones, in terms of margins or FTE’s. This, however, is not the case; instead, it appears that the opposite (weakly) holds. This means that the TCE / ex post re-negotiation approach can not be construed as an alternative explanation for our results.

9 Likewise, we have not considered another important dynamic element: The role of technological change and how integration relates to it. Afuah (2001), for instance, recently demonstrated that integration in old technologies is detrimental to overall performance, possibly because it impedes the systemic improvement of capabilities along the value chain, a theme also discussed by Chesborough (2003) and Jacobides and Winter (2005). An additional motivation to integrate, discussed by Cacciatori and Jacobides (2005), consists of the desire to fend off commoditization by providing “all-in-one” packaged services.
### TABLE 1

**What Drives Vertical Scope? Comparing the Different Theoretical Perspectives**

<table>
<thead>
<tr>
<th>Focus of the theory</th>
<th>TCE – “Measurement Branch” (<em>ex ante</em> problems)</th>
<th>TCE- Williamsonian Opportunism (<em>ex post</em> problems)</th>
<th>Capabilities and RBV Approach</th>
<th>Comparative Advantage Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes of the transaction</td>
<td>Information asymmetry and impactedness; inability to articulate what is needed</td>
<td>Potential for hold-up due to dedicated assets leads to TC</td>
<td>No focus / Prediction</td>
<td>TC is a “tax”, driven by information asymmetry, impactedness or transactional risks</td>
</tr>
</tbody>
</table>
| Attributes of the transactors | No focus / Prediction | No focus / Prediction | (a) Focus on path dependent *governance capabilities*  
(b) Focus on absolute advantage or existence of unique, rare resource to be exploited | Focus on *productive capabilities* and their distribution within segment and along the value chain (which determine total gains from trade) |
| When do firms integrate? | Information problems (due to misrepresentation or info asymmetry or inability to coordinate through prices) make integration desirable | When the threat of opportunistic renegotiation (due to asset specificity and uncertainty) is greater than the cost of internal governance | (a) The better the governance capabilities, the more the market can be used  
(b) Firms will integrate where they have absolute advantage *or* to protect a resource | When the costs of using the market (the TC “tax”) is greater than the potential *gains from trade* from specialization between any two firms |
TABLE 2  
Correlation Matrix of Key Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration in Origination</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Assets</td>
<td>-0.1736</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asset Square</td>
<td>-0.1010</td>
<td>0.8809</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% FVA loans</td>
<td>0.0596</td>
<td>-0.0748</td>
<td>-0.0349</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% FRM loans</td>
<td>-0.0664</td>
<td>-0.0173</td>
<td>-0.0172</td>
<td>-0.4304</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% ARM loans</td>
<td>0.0126</td>
<td>0.0328</td>
<td>0.0096</td>
<td>-0.2690</td>
<td>-0.1083</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTE in Production per loan</td>
<td>-0.0600</td>
<td>0.0320</td>
<td>0.0270</td>
<td>0.0283</td>
<td>0.0521</td>
<td>-0.1129</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTE in Origination per loan</td>
<td>0.2166</td>
<td>0.0491</td>
<td>0.0420</td>
<td>0.0419</td>
<td>0.0270</td>
<td>-0.1081</td>
<td>0.8903</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warehousing Margin</td>
<td>-0.1629</td>
<td>0.3136</td>
<td>0.1484</td>
<td>-0.0418</td>
<td>-0.0274</td>
<td>0.1055</td>
<td>-0.5915</td>
<td>-0.5699</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Origination Margin</td>
<td>0.2660</td>
<td>-0.1172</td>
<td>-0.0530</td>
<td>0.0229</td>
<td>-0.1486</td>
<td>-0.0865</td>
<td>0.0952</td>
<td>0.1666</td>
<td>-0.2446</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
## TABLE 4
Integration (continuous measure) in Loan Origination as a Function of Size, Transaction Costs and Capabilities
The Dependent Variable is Retail Integration (percentage of total loans made in-house) across all models and all specifications

<table>
<thead>
<tr>
<th></th>
<th>OLS (Huber-White SE)</th>
<th>Robust Regression</th>
<th>Fixed-Effects Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MODEL 1</td>
<td>MODEL 2</td>
<td>MODEL 3</td>
</tr>
<tr>
<td><strong>Total Assets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(control)</td>
<td>-1.25e-10***</td>
<td>-1.2e-10***</td>
<td>-2.3 e-10***</td>
</tr>
<tr>
<td></td>
<td>(2.39e-11)</td>
<td>(2.38e-11)</td>
<td>(1.1 e-11)</td>
</tr>
<tr>
<td><strong>Asset Square</strong></td>
<td>8.38e-21***</td>
<td>8.4e-21***</td>
<td>1.7 e-20***</td>
</tr>
<tr>
<td>(control)</td>
<td>(2.21e-21)</td>
<td>(2.21e-21)</td>
<td>(1.2 e-21)</td>
</tr>
<tr>
<td><strong>%FHA-VA loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TC measure)</td>
<td>-0.0029</td>
<td>-0.0103</td>
<td>-0.1601***</td>
</tr>
<tr>
<td></td>
<td>(0.0774)</td>
<td>(0.0813)</td>
<td>(0.0362)</td>
</tr>
<tr>
<td><strong>%FRM loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TC measure)</td>
<td>-0.1059</td>
<td>0.1073</td>
<td>-0.1360***</td>
</tr>
<tr>
<td></td>
<td>(0.0671)</td>
<td>(0.0779)</td>
<td>(0.0334)</td>
</tr>
<tr>
<td><strong>%ARM loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TC measure)</td>
<td>0.0176</td>
<td>0.1300</td>
<td>-0.041</td>
</tr>
<tr>
<td></td>
<td>(0.0993)</td>
<td>(0.0974)</td>
<td>(0.0539)</td>
</tr>
<tr>
<td><strong>FTE per loan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>origination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(productive capability)</td>
<td>-0.1629***</td>
<td>-0.1671***</td>
<td>-0.1671***</td>
</tr>
<tr>
<td></td>
<td>(0.0148)</td>
<td>(0.0088)</td>
<td>(0.0088)</td>
</tr>
<tr>
<td><strong>FTE in warehousing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(productive capability)</td>
<td>0.2349***</td>
<td>0.2551***</td>
<td>0.2551***</td>
</tr>
<tr>
<td></td>
<td>(0.0132)</td>
<td>(0.0078)</td>
<td>(0.0078)</td>
</tr>
<tr>
<td><strong>Warehousing margin</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(productive capability)</td>
<td>-0.0251***</td>
<td>-0.0239***</td>
<td>-0.0239***</td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td>(0.0035)</td>
<td>(0.0035)</td>
</tr>
<tr>
<td><strong>Loan origination</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>margin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0071)</td>
<td>(0.0062)</td>
<td>(0.0042)</td>
</tr>
<tr>
<td><strong>CONTROLS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YEAR DUMMIES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.747***</td>
<td>0.8132***</td>
<td>1.4514***</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.0592)</td>
<td>(0.1686)</td>
</tr>
<tr>
<td><strong>R^2 (%)</strong></td>
<td>5.13</td>
<td>5.69</td>
<td>50.54</td>
</tr>
</tbody>
</table>

| N                   | 1792                | 1790              | 1166                     | 1792                 | 1790             | 1166                    | 1792                 | 1790                 |

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

(Standard Errors in parentheses)
# TABLE 4

**Logit Analysis of Integration as a Function of Size, TC and Capabilities**

Dependent variable: Retail Integration (0 if <10%, 1 if >90%, dropped otherwise)

<table>
<thead>
<tr>
<th></th>
<th>MODEL 1</th>
<th>MODEL 2</th>
<th>MODEL 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Assets</td>
<td>-8.84E-10*</td>
<td>-9.32E-10*</td>
<td>-2.4E-09*</td>
</tr>
<tr>
<td>(control)</td>
<td>(4.577E-10)</td>
<td>(4.819E-10)</td>
<td>(1.04E-09)</td>
</tr>
<tr>
<td>Asset Square</td>
<td>1.026E-19</td>
<td>1.114E-19</td>
<td>3.76E-19*</td>
</tr>
<tr>
<td>(control)</td>
<td>(1.027E-19)</td>
<td>(1.062E-19)</td>
<td>(1.83E-19)</td>
</tr>
<tr>
<td>%FHA-VA loans</td>
<td></td>
<td>0.9255</td>
<td>6.2617***</td>
</tr>
<tr>
<td>(TC measure)</td>
<td></td>
<td>(0.8086)</td>
<td>(1.4691)</td>
</tr>
<tr>
<td>%FRM loans in</td>
<td>-1.1311**</td>
<td>5.0129***</td>
<td></td>
</tr>
<tr>
<td>(TC measure)</td>
<td>(0.4838)</td>
<td>(1.3193)</td>
<td></td>
</tr>
<tr>
<td>%ARM loans in</td>
<td>-0.4323</td>
<td>6.1867*</td>
<td></td>
</tr>
<tr>
<td>(TC measure)</td>
<td>(0.9408)</td>
<td>(2.9573)</td>
<td></td>
</tr>
<tr>
<td>FTE per loan in</td>
<td>-3.4460***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>origination</td>
<td>(0.5866)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(productive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>capability)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTE in warehousing</td>
<td></td>
<td>3.4576***</td>
<td></td>
</tr>
<tr>
<td>(productive</td>
<td></td>
<td>(0.5677)</td>
<td></td>
</tr>
<tr>
<td>capability)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warehousing margin</td>
<td></td>
<td>0.0556</td>
<td></td>
</tr>
<tr>
<td>(productive</td>
<td></td>
<td>(0.1503)</td>
<td></td>
</tr>
<tr>
<td>capability)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan origination margin</td>
<td>0.0893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(productive</td>
<td></td>
<td>(0.1817)</td>
<td></td>
</tr>
<tr>
<td>capability)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.1040***</td>
<td>2.5813***</td>
<td>-1.7352</td>
</tr>
<tr>
<td></td>
<td>(0.1672)</td>
<td>(0.3693)</td>
<td>(1.9987)</td>
</tr>
<tr>
<td><strong>Pseudo R^2 (%)</strong></td>
<td><strong>1.61</strong></td>
<td><strong>4.21</strong></td>
<td><strong>72.62</strong></td>
</tr>
<tr>
<td>LR Chi2</td>
<td>13.31</td>
<td>34.68</td>
<td>276.21</td>
</tr>
<tr>
<td>N</td>
<td>1106</td>
<td>1104</td>
<td>643</td>
</tr>
</tbody>
</table>

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001 (Standard Errors in parentheses)

Logit Analysis performed with robust standard errors, clustered on Firm ID to account for non-independence of errors. Likelihood Ratio Chi^2 is reported from the non-robust model.
FIGURE 1
The Value Chain in Mortgage Banking

Figure 1A: The Broader Mortgage Banking Value Chain

Market for Closed Loans → Market for loan bundles → Secondary loan market

Origination → Warehousing → Securitizing and Payment processing → Holding loan Prepayment risk → Servicing

Mortgage Brokers or Mortgage Banks → Mortgage Banks → GSE’s and securitizers → Wall Street Players → Mortgage Banks

Figure 1B: The Activities in the Mortgage Bank’s Value Chain

Loan Origination → Warehouse/Pipeline Analysis → Transfer to Secondary Market

Pre-Qualification - Document Generation - Application Processing - Credit Analysis & Underwriting - Approval & Closing

- Loan Limit Estimates
- Loan Structuring Options
- Maximum Monthly Payment Estimates
- Application Documents
- Disclosure Documents
- Compliance Documents
- Credit Analysis Worksheets
- Document Tracking
- Payment Calculations
- Status Monitoring
- Escrow Structuring
- Appraisal
- Title Search
- Credit Checking and Scoring
- Closing Calculations
- Closing Documents
- Setup for Servicing

Loan Servicing

- Payment Processing & Reporting
- Escrow Management
- Customer Service
- Collections, Bankruptcies & Foreclosures

- Payment Accounting
- Statements
- Tax Reporting
- Hazard Insurance
- Property Tax Accounting
- Balance Inquiries
- Escrow Inquiries
- Late Payment Notices
- Management of Delinquent Accounts

Note: The activity which mortgage banks can “make” themselves or “buy” through the market is Loan (i.e. retail) origination. This activity, with these five value-adding steps, is our focus.
Appendix 1:

Measuring Transaction Risk Using Loan Types

We perform three sets of tests to validate the observation that loan type is a surrogate for risk.

First, using robust regression, we examined the relationship between integration and overall default risk. Our results suggest that default risk is indeed correlated (significant at p<.05) with market procurement. From the regression estimates, we calculate that an entirely integrated bank faces a default rate of 1.72%, while the rate for a bank that buys all loans on the market is 2.06%.

Second, we compare default and delinquency rates of each loan category for each firm over time using the non-parametric Wilcoxon test. Our results suggest that FHA/VA loans have default rates substantially higher than ARM or FRM loans (p<.001). FRMs also have higher default rates than ARMs but the contrast is less substantial (p<.1). We obtain similar results using an alternative data source for overall defaults and delinquencies by loan type from the Department of Housing and Urban Development (HUD) over our sample period suggesting the patterns in our data are consistent with the industry-wide demographics in delinquencies and defaults. However, the fact that some loans are riskier is not enough to establish the existence of differential transaction costs.

This moves us to the third set of tests. Specifically, to use these loan categories as surrogate measures for the extent of transaction risks, we also need to demonstrate that riskier loans are comparatively riskier to procure through the market; that is, that the riskier the loan, the more beneficial it is to produce it in-house. This leads to our final analysis, in which we use robust regression to compute correlations between the use of retail production (integration) and the percentage of delinquent loans by loan type. We find that the more a bank is integrated, the lower the risks of default losses, but, crucially, that the correlation patterns vary depending on the loan type. The correlation is highly significant for the more dangerous FHA/VA loans – regression estimates suggest that a fully integrated firm faces losses on these loans of about 1.43% in contrast to a 2.77% loss rate for a specialized firm (significant at p<.001). This difference drops to .16% for FRMs (p<.05) and is not significant for ARMs. So for the “risky” loans, the use of the market leads to significant costs of misrepresentation; whereas for the standard loans, there is little, if any, transactional hazard induced by using the market. This means that loan types do provide a reasonable direct measure of the transactional hazards a firm faces, at least with regard to misrepresentation and measurement (Akerlof, 1970).