LOSING SIGHT OF THE FOREST FOR THE TREES? 
PRODUCTIVE CAPABILITIES AND GAINS FROM 
TRADE AS DRIVERS OF VERTICAL SCOPE

MICHAEL G. JACOBIDES1* and LORIN M. HITT2
1 London Business School, London, U.K.

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 can shape vertical scope through gains from trade. Using highly detailed data from 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.

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 will cause firms to increase vertical scope, while the absence of such hazards and gains from specialization cause firms to use ‘the market.’ Numerous empirical studies on vertical scope have been produced over the last 15 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 panoply of 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 adds a focus 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 ‘comparative advantage’ (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 (Ricardo, 1817): 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. As with international trade, sufficiently high gains from contracting can often overcome procurement costs and cause firms to specialize vertically. Conversely, if productive capabilities do not vary much across firms, then vertical specialization will not occur even if the net TC ‘tax’ is extremely low, because there is little potential gain from such contracting. In this regard, our perspective is analogous to the resource-based view (RBV) 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 basis for specialization at the firm or industry level.

Our approach differs somewhat from recent applications of the resource-based view, 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; 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 the mortgage industry, focusing on productive capabilities (for greater detail, see Jacobides, 2004, and 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 both owned 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. Several characteristics of mortgage banking make the industry especially well suited to empirical research of this kind. The industry’s use of clearly and uniformly defined products and the value chain segments enables us 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, it is the productive capabilities view that 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 in which productivity differs substantially across different 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 these drivers of vertical scope, this section summarizes the motivations for vertical integration, 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 that the greater the hazards of market exchange, the greater the tendency for firms to produce intermediate products internally rather than purchase them on the external market. Such hazards arise due to either 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 in which market transactions can be costly or impossible due to the 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 from 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, Meehan, and Snyder, 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. This information asymmetry creates 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) provide evidence on such measurement costs as they affect both the choice of vertical scope and the resulting satisfaction with the governance choice.

Uncertainties in the business environment may cause firms to avoid 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). Since relationship-specific investment is often essential to efficient production, these risks can reduce the likelihood of trade. This perspective has been extensively researched and documented (Shleifski and Klein, 1995; Williamson, 1999), 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. Information technology, especially automated credit scoring and underwriting, have further reduced asset specificity (see Jacobides, 2005). Therefore, 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:

**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 subsection summarizes the relevant background and related literature and then briefly describes our perspective on the effects of comparative advantage 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?’ (Williamson, 1999: 1103), a theme also pursued by Madhok (2002). On this broader view of firm scope, there has been significantly less empirical evidence.

Whereas 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...
capabilities have not played a substantial role in the empirical literature until recently. Argyres (1996) was one of the first to provide qualitative evidence on the role of firm capabilities, observing that business capabilities were a significant driver of vertical scope in the cable manufacturing industry. Work on the resource-based view of the firm further led scholars to consider the role of firm heterogeneity. Skill sets (Poppo and Zenger, 1998), specific experience and expertise (Leiblein and Miller, 2003; Hoetker, 2005), and contracting ability (Argyres and Liebeskind, 1999; Madhok, 2002) have been shown to be important determinants of outsourcing decisions.

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 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, with its internal cost levels, vs. the price and transaction costs of interacting with another firm (Jacobides and Winter, 2005).

To make headway in a theory of capability-based analysis of vertical scope, we must further refine the concept of capability. An important distinction we introduce in this paper is the distinction of productive capabilities of a firm (productive efficiency or ‘zero-order capabilities’—see Winter, 2003) in each of the vertically related stages, vs. the capabilities of governance, the ability of a particular firm to use integration or the market to create value by linking these 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 treated relative capability (\( p_i - p_j \)) and internal governance costs \( B_i \) as constants (Williamson, 1985: Ch. 6) and treated transaction costs as an industry characteristic. The more recent TCE literature has, in addition, allowed for firm-specific variation in transaction costs (\( TC_{ij} \)) due to efficiencies in contracting with specific suppliers (j) (see Dyer, 1996) or buyers (i) (Winter, 1988; Argyres and Liebeskind, 1999; Leiblein and Miller, 2003; Mayer and Argyres, 2004). Thus, the emphasis of recent TCE research is 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 which firms face in their operations (Levinthal, 1997). To the extent that process knowledge is sufficiently complex to limit imitation (Porter, 1996; Rivkin, 2001; Siggelkow, 2001) or include irreversible investments that alter future opportunities (Ghemawat, 1991; Winter, 1995), 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 et al., 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
Losing Sight of the Forest for the Trees?

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 with strong capabilities upstream and downstream will be integrated because it can avoid the inefficiency of their downstream activity, and provide additional capacity to expand their efficient upstream operations. Thus, variations in capabilities can lead to variations in governance choices. This observation leads to two specific predictions about vertical scope. First, the more capable a firm is in a particular segment, the more that firm will participate in that segment. Thus:

**Hypothesis 2:** Superior productive capability in a particular vertical segment will be positively associated with activity in that same vertical segment.

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 \((p^{U_i} > p^{U})\), the smaller the probability that it is vertically integrated into the downstream segment, all else being equal. Firms with strong upstream capabilities will want to use the intermediate market to become a net seller of upstream goods, and, as such, the firms 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 \(p^{U_i} > p^{U}\). 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 provides an explanation for why firms may be active in both vertical segments, consistent with the fact that 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 segment to grow more in its area of comparative advantage. The existence of sunk investments and limits to the speed of entry or exit in a segment underpin Hypotheses 2 and 3, and can explain both ‘mixed governance’ and the gradual changes in firms’ scope often observed in practice.¹

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.

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 that wants to grow and leverage its absolute competitive advantage. Because 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 (see Jacobides, 2004).

Copyright © 2005 John Wiley & Sons, Ltd.

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, 2002). The variation in the ownership of these assets causes variation in integration, depending on the relative costs of acquiring and integrating the requisite complementary resources and capabilities (Barney, 1999). Although 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-lived comparative advantage—may explain a significant part of the variance in the way firms choose their vertical scope.2

To recap, different theories have shed light on what drives vertical scope. Table 1 summarizes the major approaches, identifying their focus and predictions for where we expect to see integration, as a function of the attributes of the transaction (information asymmetry; threat for opportunism in a renegotiation 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 considers the relationship between integration and the factors driving integration as predicted by the measurement branch of transaction cost economics (Hypothesis 1) and our ‘comparative advantage’ perspective (Hypotheses 2 and 3). We are not only interested in testing these hypotheses but also in 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 provides a good setting to study vertical scope because it has a complex value chain which has been increasingly fragmented into quasi-independent parts, and it includes both integrated firms as well as narrowly targeted specialists (Jacobides, 2005). The structure of the mortgage banking industry is presented in Figure 1(a), which shows the vertically co-specialized world of mortgage banking (which itself competes with more integrated savings and loan institutions and banks that also offer mortgages). The industry has 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 percent 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. Thus our dependent variable, the degree of integration, represents a true firm choice. However, although there is no formal regulatory framework for the industry, mortgage banks are called upon by their association and by 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 1(b). 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.

2 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 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.
Table 1. What drives vertical scope? Comparing the different theoretical perspectives

<table>
<thead>
<tr>
<th>Focus of the theory</th>
<th>TCE—‘measurement branch’ (ex ante problems)</th>
<th>TCE—Williamsonian opportunism (ex post 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>
<tr>
<td>Attributes of the transactors</td>
<td>No focus/Prediction</td>
<td>No focus/Prediction</td>
<td>(a) Focus on path-dependent governance capabilities (b) Focus on absolute advantage or existence of unique, rare resource to be exploited</td>
<td>Focus on productive capabilities and their distribution within segment and along the value chain (which determine total gains from trade)</td>
</tr>
<tr>
<td>When do firms integrate?</td>
<td>Information problems (due to misrepresentation or info asymmetry or inability to coordinate through prices) make integration desirable</td>
<td>When the threat of opportunistic renegotiation (due to asset specificity and uncertainty) is greater than the cost of internal governance</td>
<td>(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</td>
<td>When the costs of using the market (the TC ‘tax’) is greater than the potential gains from trade from specialization between any two firms</td>
</tr>
</tbody>
</table>
Figure 1. The value chain in mortgage banking. 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.
Table 2. Correlation matrix of key variables

<table>
<thead>
<tr>
<th></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</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>origination</td>
<td>A</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total assets</td>
<td>B</td>
<td>−0.1736</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asset square</td>
<td>C</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>
</tr>
<tr>
<td>% FVA loans</td>
<td>D</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>
</tr>
<tr>
<td>% FRM loans</td>
<td>E</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>
</tr>
<tr>
<td>% ARM loans</td>
<td>F</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>
</tr>
<tr>
<td>FTE in</td>
<td>G</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>
</tr>
<tr>
<td>production per loan</td>
<td>H</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>
</tr>
<tr>
<td>Warehousing</td>
<td>I</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>
</tr>
<tr>
<td>margin</td>
<td>J</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>
</tr>
</tbody>
</table>

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 1(b) 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 Figure 1(b): pre-qualification, document generation, application processing, credit analysis and underwriting, and approval. Productive capabilities in warehousing affect the performance of 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, and productivity metrics enabling the segments to be evaluated independently.

Data: Mortgage banks’ integration into retail loan production

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 and avoiding potential self-selection biases. 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 was sampled once every 3.1 years, on average, during the 10-year period of our dataset, and the total number of firms in the sample is 685. On average, the respondents of the MBFRF are responsible for about 25 percent of the total loans produced in the United States; as such, this is a highly representative sample. The database we utilize covers the period 1988–98 and contains 1792 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

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 (both buying and making loans). We take the perspective of all banks that warehouse a loan, and we consider whether these banks make their own loans, buy their loans, or do both. Thus, 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.3

Transaction costs

From our industry interviews we determined that the principal risks of using the market for the procurement of loans were the risks of trading loans of unknown quality. Borrowers can vary in ways that are difficult to observe or objectively measure, and those variations influence the borrower’s probability of delinquency or default—a very real cost to the mortgage banks.4 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 loan 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 the heterogeneity of the eligible pool of borrowers, and the fact that their credit standards are more generous as a matter of social policy. ‘Plain vanilla’ fixed-rate conventional mortgages (FRMs) have intermediate levels of risk due to their long-term and (relatively) high interest rate, and short-term, low-rate adjustable rate mortgages (ARMs) are the least risky, as industry-wide default rates suggest.

To confirm that transaction risks follow this categorization, we conducted several tests (see 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 represents a direct measure of transaction risks relevant to 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 vs. 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 productivity5 and industry practice (Posner and Nambiar, 2002). Because outputs are relatively standardized in this industry and the MBFRF takes great care to properly capture revenue, staffing and total cost for each segment,

---

3 Most banks in our sample use the mixed mode, so that the degree of integration varies continuously from 0 percent to 100 percent specialized in origination. This continuous, measurable variation in integration provides a significant measurement advantage because 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.

4 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. Second, 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.

5 We utilize labor 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 because 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, however, because the MBFRF does not include an inventory of non-financial capital.
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 by the origination segment or serviced in the warehousing segment) per full-time equivalent employee (FTE) in each segment. Results are similar whether we use the number of loans or the dollar volume of loans. Operating margins are obtained in each segment through the MBFRF income breakdown which isolates segment-specific cost and revenue. Whereas 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 (Hypothesis 2) and negatively related to efficiency in warehousing (Hypothesis 3).

Control variables
We include control variables for time (a dummy variable for the year) 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, because their own branches offer insufficient capacity to service 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 because assets and returns can vary more than in ‘industrial’ corporations. This is particularly useful because 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, because during the observation period there has been substantial merger and acquisition activity; 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 sample 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 cut-off points, including 0 percent and 100 percent.

In all four methods, we sequentially introduce explanatory variables for scale, for transaction cost-related factors, and finally for 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
Table 3. 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>Asset square (control)</td>
<td>8.38E–21***</td>
<td>8.4E–21***</td>
<td>1.88E–21</td>
</tr>
<tr>
<td>(control)</td>
<td>(2.21E–21)</td>
<td>(2.21E–21)</td>
<td>(2.61E–21)</td>
</tr>
<tr>
<td>% FHA–VA loans (TC measure)</td>
<td>-0.0029</td>
<td>-0.0103</td>
<td>0.001601***</td>
</tr>
<tr>
<td></td>
<td>(0.0774)</td>
<td>(0.0813)</td>
<td>(0.0362)</td>
</tr>
<tr>
<td>% FRM loans (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>% ARM loans (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>(0.0621)</td>
</tr>
<tr>
<td>FTE per loan, origination (productive capability)</td>
<td>-0.1629***</td>
<td>-0.1671***</td>
<td>-0.1107***</td>
</tr>
<tr>
<td></td>
<td>(0.0148)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTE in warehousing (productive capability)</td>
<td>0.2349**</td>
<td>0.2551***</td>
<td>0.1453**</td>
</tr>
<tr>
<td></td>
<td>(0.0132)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warehousing margin (productive capability)</td>
<td>-0.0251***</td>
<td>-0.0239***</td>
<td>0.00048</td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan origination margin (productive capability)</td>
<td>0.0071</td>
<td>0.0136**</td>
<td>0.0163**</td>
</tr>
<tr>
<td></td>
<td>(0.0062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Year dummies</td>
<td>Year dummies</td>
<td>Year dummies</td>
</tr>
<tr>
<td>Constants</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>$R^2$ (%)</td>
<td>5.13</td>
<td>5.69</td>
<td>30.54</td>
</tr>
<tr>
<td>$N$</td>
<td>1792</td>
<td>1790</td>
<td>1166</td>
</tr>
</tbody>
</table>

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; **** $p < 0.001$ (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>Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets (control)</td>
<td>$-8.84E-10$ (4.577E-10)</td>
<td>$-9.32E-10$ (4.819E-10)</td>
<td>$-2.4E-09$ (1.04E-09)</td>
</tr>
<tr>
<td>Asset square (control)</td>
<td>$1.026E-19$ (1.027E-19)</td>
<td>$9.255$ (0.8086)</td>
<td>$6.261E-19$ (1.4691)</td>
</tr>
<tr>
<td>% FHA–VA loans (TC measure)</td>
<td>$-1.1311$ (0.4838)</td>
<td>$5.0129$ (1.3193)</td>
<td>$-3.4460$ (0.5866)</td>
</tr>
<tr>
<td>% FRM loans in (TC measure)</td>
<td>$3.76E-19$ (1.83E-19)</td>
<td>$2.4E-09$ (1.04E-09)</td>
<td></td>
</tr>
<tr>
<td>FTE per loan in origination (productive capability)</td>
<td>$4.12E-09$ (8.95E-09)</td>
<td>$2.4E-09$ (1.04E-09)</td>
<td></td>
</tr>
<tr>
<td>FTE in warehousing (productive capability)</td>
<td>$-1.1311$ (0.4838)</td>
<td>$5.0129$ (1.3193)</td>
<td>$-3.4460$ (0.5866)</td>
</tr>
<tr>
<td>Warehousing margin (productive capability)</td>
<td>$0.0556$ (0.1503)</td>
<td>$0.0893$ (0.1817)</td>
<td></td>
</tr>
<tr>
<td>Loan origination margin (productive capability)</td>
<td>$2.1040$ (0.1672)</td>
<td>$2.5813$ (0.3693)</td>
<td>$-1.7352$ (1.9987)</td>
</tr>
<tr>
<td>Constant</td>
<td>$2.1040$ (0.1672)</td>
<td>$2.5813$ (0.3693)</td>
<td>$-1.7352$ (1.9987)</td>
</tr>
<tr>
<td>Pseudo $R^2$ (%)</td>
<td>$1.61$</td>
<td>$4.21$</td>
<td>$72.62$</td>
</tr>
<tr>
<td>LR $\chi^2$</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.

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 a linear nor a monotonically positive function of scale. Similar results were obtained in the Logit regressions of Table 4, although they are less strongly significant.

Hypothesis 1 (on the role of TC) received mixed support, as we can see through Model 2, which includes controls and the TC-related measures. In the logistic regression—which considers what variables explain firms’ choices to be entirely integrated or entirely specialized—the sign and magnitude of the coefficients are generally what the theory would predict (see Table 4). Relatively safe loans (FRM) are related negatively to the decision to be fully integrated. 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 vs. 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, the robust and fixed effects regressions yield results that occasionally contradict the transactional logic.8

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 percent (from 1.6% to 4.2%). To put this in context, the addition of capability metrics increases the pseudo-$R^2$ to 72.6 percent (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 percent to 5.7 percent in OLS, and the within-$R^2$ in fixed effects increases from 4.2 percent to 5.7 percent.

8The 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 because 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).
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. The capability-related coefficients in terms of productivity (FTE per loan) are significant in the expected direction. Thus, banks that are more efficient downstream (low FTE/loan) show a greater extent of integration into that segment. Firms that are better than others in origination are integrated in that vertical segment (Hypothesis 2). Furthermore, the coefficient of the efficiency in the upstream part of production (loan warehousing) is strongly negatively associated with integration (Hypothesis 3). Similar findings are obtained 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 reliance on the market (Hypothesis 3). Again, this dramatically increases explanatory power; in OLS, the $R^2$ went from 5.6 percent to 50.5 percent.

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 Hypotheses 2 and 3—most of the coefficients are highly significant in the same direction. The reduction of significance was expected in a fixed-effects specification because 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 Hypotheses 2 and 3 because they suggest that capability differences and integration not only co-vary between firms but also over time for the same firm. We are somewhat guarded about this analysis, however, because the fixed effects estimates can be contaminated by changes in the nature of a firm resulting from M&A.

The Logit analysis is also consistent with OLS, robust, and fixed effects regressions. The coefficients for both productivity and margin are significant and in the hypothesized direction, except for the warehousing margin measure (Hypothesis 3b). In addition, when we add the capability-based metrics, pseudo-$R^2$ in the Logit regression increases from 5.6 percent to 72.6 percent.

**DISCUSSION**

This analysis suggests that, at least in our particular setting, differences in productive capabilities are key drivers of the decision to integrate. Although a transactional logic may partly explain our observations, capability considerations dominate. Although this finding may relate to the specificities of our setting, it also suggests 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 confirms the need to include capability measures 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), consistent with prior literature that suggests firm-specific factors affect governance choice (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 and can lead to firm variation in governance choice and the use of ‘mixed-mode’ governance structures. Since variation in capabilities is persistent and important in a wide variety of industries (e.g., Lieberman and Dhawan, 2001) and value chains (cf. Duncan, 1998), this observation may be relevant to a substantial segment of economic activity.

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.’ Simple logic suggests that if a firm has a comparative advantage in a particular part of the value chain,
another firm has a corresponding weakness creating opportunities for gains from trade. Strong firms in one activity will trade with strong firms in other, vertically related activities. Note that our theory is based on comparative advantage along the value chain rather than ‘absolute’ advantage of the firm as a whole, distinguishing our analysis from the RBV of vertical scope decisions (Combs and Ketchen, 1999; Schilling and Steensma, 2002). We also find that scale is not a primary driver of specialization. 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 avenues 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, reinforced 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 cause us to reintegrate 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 (Cacciatore and Jacobides 2005; Coase, 1992; Jacobides and Winter, 2005; Jacobides, 2005). TC determine the extent to which latent gains from trade—driven by the distribution of productive capabilities in an industry—are or are not sufficient to justify vertical specialization. The joint analysis of differences in productive capabilities and transaction costs/vertical scope holds much promise for the future and opens up new avenues of research.

Limitations

This study is subject to a number of limitations. On the empirical level, we do not directly test Williamsonian TC. While we 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 do provide measures related to hold-up. Thus, our findings cannot rule out that further unobserved TC are empirically important. That being said, the measures we used for capability differences cannot be reasonably construed as indirectly associated with Williamsonian TC. So our basic finding—i.e., that productive capabilities explain a large part of the variance—still stands.

9 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.

10 In our setting (and, we suspect, many others) it is hard to conceive 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, but we did explore it econometrically. Specifically, we examined whether integrated firms (which by definition do
On the theoretical level, there are additional limitations. First, our paper presumes the existence of a clear, separable governance interface; that is, the paper is built on the assumption that there is a potential market, with some level of TC ‘tax.’ Although 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.

Second, while our research points out that differences in productive capabilities matter, and that the distribution of productive capabilities thus drives scope, we do not focus on why capabilities may be vastly different in some settings while very similar in others. Similarly, the emphasis of our theory is on the relationship of capability differences and scope rather than how these capabilities differences arise, themes recently discussed by Jacobides and Winter (2005), Gulati et al. (2005), or Cacciatori and Jacobides (2005).

Third, we did not directly consider any of the ‘dynamic’ drivers of integration such as protection or creation of rents from 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).\footnote{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 Chesbrough (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.}

Finally, we do not consider the relative costs of different mechanisms for developing capabilities such as organic growth or M&A (see Karim and Mitchell, 2000, or Helfat and Raubitschek, 2000). Thus, while we can suggest such capability development should lead to a reconsideration of governance choice, we cannot address the managerial question of when it is worthwhile to develop capabilities to better exploit gains from trade. We consider these topics as important extensions of our research.

**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.
ACKNOWLEDGEMENTS

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, Sumanta Ghoshal, Costas Markides, and Margarethe Wiersema for useful comments; and Srikanth Kannan for research assistance. We would also like to thank Associate Editor 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.

REFERENCES


APPENDIX: 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 \( p < 0.05 \)) with market procurement. From the regression estimates, we calculate that an entirely integrated bank faces a default rate of 1.72 percent, while the rate for a bank that buys all loans on the market is 2.06 percent.

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 \( (p < 0.001) \). FRMs also have higher default rates than ARMs but the contrast is less substantial \( (p < 0.1) \). 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 percent in contrast to a 2.77 percent loss rate for a specialized firm \( (p < 0.001) \). This difference drops to 0.16 percent for FRMs \( (p < 0.05) \) and is not significant for ARMs. So for the ‘risky’ loans, the use of the market causes 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).