Comparison of long-term contracts and vertical integration in decentralised electricity markets

Richard Meade
Seini O’Connor
Abstract: A common view is that electricity sectors should be unbundled and opened to competition, and that in such decentralised markets long-term contracts are necessary to constrain generator market power, elicit competitive entry in retailing, and support new generation investment and hence supply security. However, increasing levels of generation ownership by electricity retailers or large electricity customers (and vice versa) – i.e. of vertical (re-)integration - have been observed in several liberalised systems. In this paper, we examine problems arising in contract markets in electricity systems. The resulting shortcomings in electricity contract markets suggest they are at best a partial and unsustainable solution to market power, retail competition, investment and supply security issues. We argue that high levels of vertical integration should not be a cause for concern, but rather may represent a more “natural” structure for the electricity sector. Vertical integration is argued to be a more self-sustaining institutional arrangement, and one which better addresses issues of wholesale market power, investment, and supply security. Its endogenous rise – even in electricity systems with relatively liquid contract markets – further suggests it has a natural role to play in decentralised electricity systems. Arguments are presented that vertical integration may in fact become a necessary form of industry organisation once a threshold level of integration has been reached.

1 New Zealand Institute for the Study of Competition and Regulation (R. Meade, richard.meade@cognitus.co.nz, S. O'Connor, seini.oconnor@vuw.ac.nz). This paper draws on a paper of the same name presented at the “Efficiency, Competition and Long Term Contracts in Electricity Markets” workshop organised by LARSEN and the Loyola de Palacio Chair of the European University Institute, European University Institute, Florence, 15-16 January 2009. It has benefited from comments by discussants at that workshop, and by those at presentations of the paper given subsequently at Victoria University of Wellington and the University of Canterbury. As usual, all errors or omissions are the responsibility of the authors.
1. Introduction

A common view is that electricity sectors should be unbundled and opened to competition, and that in such decentralised (liberalised) markets long-term contracts are necessary to constrain generator market power, elicit competitive entry in retailing, and support new generation investment and hence supply security (Boom & Buehler (2006), Vázquez et al. (2002)). However, increasing levels of generation ownership by electricity retailers or large electricity customers (and vice versa) – i.e. of vertical (re-)integration - have been observed in several liberalised systems (Thomas (2004), Anderson et al. (2007), Hogan & Meade (2007); Gans & Wolak (2008)). This has led to concerns about market performance, with policy-makers concerned that wholesale and retail competition might decrease, and generation investment and supply security may be under threat (European Commission (2007); Michaels (2006)).

However, problems have emerged with this “conventional” view regarding sector unbundling and competition (Green, 2004; Chao et al., 2005, 2008; Finon & Perez, 2008). In particular, excessive competitive entry in retailing may threaten the viability of short- and long-term contracts, and thus threaten—rather than promote—generation investment and supply security. Excessive competition can take the form of “hit-and-run” retail entry, for example, under which retail entrants with low entry costs predate customers from existing retailers when the incumbents have signed long-term energy supply contracts at fixed prices but then wholesale energy prices fall. This creates critical “hold-up” risks between retailers and consumers – i.e. the risk that customers, facilitated by low switching costs and short contract durations, abandon their contracts with retailers, leaving them with contracts they then cannot honour. In turn this translates into cascading hold-up risks for parties having upstream contracts with those retailers, notably generators (and, ultimately, their fuel suppliers and financiers). Ironically, this leads to a conundrum – a liquid contracts market is seen as necessary to elicit retail entry, but risk of too much entry undermines the supply and viability of such contracts.

In this paper, we examine problems arising in contract markets in electricity systems, identifying possible causes of hold-up as well as other contracting problems that can result in inadequate investment and supply security, and increased threat of market power. Specifically, hold-up risks arise not just from the threat of competitive retail entry, but also due to the competition faced by industrial customers, and other factors such as generator fuel mix and fuel supply security. Additionally, contracting in electricity systems is further complicated by factors such as mismatches in the preferred load profiles of generators and their customers, information asymmetries regarding generator outage rates and fuel security, generator market power, and strategic bargaining (in which more informed parties negotiate better contractual terms than their less informed counter-parties). The resulting shortcomings in electricity contract markets suggest they are at best a partial and unsustainable solution to market power, retail competition, investment and supply security issues. Instead we argue they are largely a disequilibrium institution, often arising artificially.

1 In this paper we set aside questions of whether “natural monopoly” network activities such as high voltage transmission or low voltage distribution should or should not be integrated with potentially competitive activities such as generation and energy retailing. We assume throughout that network activities are not combined with generation or retailing, but see Littlechild (2008) and Meade (2005) for discussions of examples where they might be efficiently combined.
as an imposed element of electricity liberalisations reflecting initial reform priorities and concerns.

We argue that high levels of vertical integration should not be a cause for concern, but rather may represent a more “natural” structure for the electricity sector. Where contracts have failed to emerge or operate as expected, and political priorities have shifted away from constraining generator market power in favour of ensuring supply security, the relative importance of contracts in the scheme of electricity system governance has diminished in favour of increased vertical integration (with contracting and spot energy markets providing complementary roles). Such vertical integration is, by contrast, argued to be a more self-sustaining institutional arrangement, and one which better addresses issues of wholesale market power, investment, and supply security. Its endogenous rise – even in electricity systems with relatively liquid contract markets – further suggests it has a natural role to play in decentralised electricity systems. Arguments are presented that vertical integration may in fact become a necessary form of industry organisation once a threshold level of integration has been reached.

We begin in Section 2 by discussing the expectations and experience of contracting in decentralised electricity systems, noting rising levels of vertical (re-)integration and discussing research on the impact of this trend. Section 3 then sets out our analytical framework, drawing on the transaction cost economics and property rights literatures, and empirical literatures on both contracting and vertical integration. We stress the importance of contracts in addressing hold-up problems, but also the circumstances in which ownership (specifically, vertical integration) is preferable. Section 4 examines in detail the particular problems confronting contracting in electricity systems, and the resulting shortcomings of contracting. Section 5 then sets out the ways in which vertical integration better addresses those problems than does contracting. While vertical integration is argued to be a more natural and self-sustaining primary approach to addressing electricity sector problems – such as risk management, securing investment, addressing generator market power, and supporting retail entry – it is argued that both spot and long-term market contracting remain natural complements to vertical integration. Alternative solutions to the problems of contracting in electricity are also discussed. Finally, Section 6 discusses the resulting policy implications. We argue for a variation on proposals that have been made for the retention or reinstatement of retail franchise monopolies as a means to constrain hold-up risks between retailers and consumers, with some intermediate level of competition – between nil and absolute – being optimal. Hence, if policy responses are required to ensure supply security while also eliciting workable but not excessive retail competition, they should include less focus on generator market power, a greater tolerance of vertical integration, and a reduced emphasis on contracting.

2. The expectations and experience of contracting and the re-emergence of vertical integration

In this section we set out the expectations and experience of contracting in a variety of liberalised electricity systems, noting the shortcomings revealed in the “conventional” liberalisation model’s assumptions about the role of contracting. We observe the increasing levels of vertical (re-)integration emerging in many sectors in response to these shortcomings, and briefly refer to evidence on the implications of this trend for sector performance.
2.1. The expectations of contracting

The “conventional” model of electricity sector liberalisation holds that vertical and horizontal integration across all parts of the system is not a necessary form of organisation. In particular, while transmission and distribution have “natural monopoly” features that limit their contestability, it is possible to introduce welfare-enhancing competition across generation and retailing. This requires ensuring that existing and entrant generators and retailers have non-discriminatory access to network assets. Some form of wholesale market is also needed to facilitate transparent and competitive price discovery. Such markets include centralised pools, bilateral trading, energy-only spot markets, and a range of real time and forward (including day-ahead) energy and capacity markets.

As noted by Chao, Oren & Wilson (2005), and Finon & Perez (2008), proponents of this model made two important assumptions about the efficacy of vertically separating generation from retailing, and of market-based competition:

- that contracts markets, complemented by spot markets, would naturally develop or could be sustainably imposed to replace the formerly integrated arrangements, and
- that generators would be able to raise capital to fund new investments without the security of cost recovery (whether via cost of service regulation of private utilities as in the US, or through consumer franchises or taxpayer guarantees in state-owned systems such as in the UK, Australia and New Zealand).

As to the first assumption, of particular importance is contracting between generators and either large customers or energy retailers that aggregate portfolios of smaller customers. Assuming liquid physical and financial markets for such contracts are developed over a suitable range of contract horizons and contract types, then a number of benefits could be expected to flow:

- Large customers should be able to hedge their electricity price risks through direct purchases of wholesale contracts tailored to match their load profile, and rely on their scale to counter-veil against any market power held by oligopolistic generators (see discussion in Chao et al., 2005; Anderson, Hu & Winchester, 2007; Finon & Perez, 2008). Risk management in turn is important for facilitating financing and investment, which has implications for supply security, and business survival.1
- Retailers should be able to reduce their exposure to demand uncertainty by pooling numerous smaller customer loads, and run a portfolio of supply contracts to match the risk characteristics and profile of that pooled load.
- A liquid contracts market, supported by institutional changes to ensure retail customer contestability (e.g. switching rules), should enable the competitive entry of retailers, thereby ensuring competitive pressure is placed on retail margins.
- The ability of entrant generation to trade through a spot market or other contracts markets should introduce the prospect of increased competition in wholesale prices.

1 Risk management by electricity operators is also important at a political level, in terms of supply security and price stability, since failures to achieve any of these increases the risk of large customer or voter backlash.
• Generators making new investments should be able to secure their investment returns by entering into suitable long-term contracts with one or more parties, which should go some way towards satisfying the second of the key assumptions underlying liberalisation.

However, experience in many countries has not lived up to these predictions.

2.2. Experience with contracting

Most notably, liberalisation has not led to as much competition in wholesale and retail markets, as many contractual forms to support long-term investment, or as favourable sector outcomes (in terms of generation adequacy and lower prices) as expected. In contrast, vertical integration has increasingly arisen, where it has been permitted, of its own accord.

Retail entry and generation investment. Despite entry by start-up independent electricity retailers, diversifying gas retailers, and incumbent electricity retailers from other regions, retail entry in unbundled sectors has been lower than expected (Defeuilley, 2009). For example, even at their peak in 1999-2001, independent retailers were unable to secure more than a two percent market share in the UK (which along with Norway and Sweden was one of the more vibrant retail markets) because they were unable to secure adequate physical and financial contracts to hedge their exposure to wholesale electricity prices. The remaining independent firms have integrated upstream with generation to survive, with the exception of some horizontally-integrated firms—for example, Centrica, the former gas utility, is now the only major electricity retailer in the UK without generation interests.

Similarly, generation investment in these sectors has been lower than expected. Although significant investment arose in some markets when generation was first opened to competition (e.g. in the US—Joskow, 2006), investors appear to have become more wary, and long-term contracts have not developed or performed to the satisfaction of project financiers (e.g. see de Luze, 2003). In the 1980s independent power producers were able to project finance new generation (typically gas-fired CCGT) using high debt levels on the strength of supply contracts written with utilities that retained some measure of customer franchise. In the 1990s such project financing of merchant generation continued, but without contracting. Instead, new entrants relied on high wholesale prices to finance investments—to the detriment of many when wholesale prices fell and gas prices rose. The resulting bankruptcies and withdrawal of merchant generation in the US (notably California in 2001) has been reflected in similar experiences in the UK and elsewhere in Europe.

Contract markets. More in line with predictions, contract markets have developed to a significant degree in a number of electricity systems (Chao et al., 2005; Anderson et al., 2006; Finon & Perez, 2008). Among these are the UK and Australia’s east coast National Electricity Market (NEM). In both cases generation was unbundled or privatised (or both) with long-term vesting contracts in place to smooth the transition to competitive markets, providing both price certainty and constraints on remaining generator market power. The UK also radically restructured the centralised wholesale “pool” in England and Wales and replaced it with NETA in 2002, which relies almost exclusively on bilateral contracting with only a limited centralised real time balancing market. By contrast, contract markets developed organically in Scandinavia’s NordPool and in Germany, driven by grid companies and utilities respectively.
Where contract markets have developed, however, they have not developed to the extent envisaged. Anderson et al. (2006) report that base load generators in the NEM are 70-80 percent hedged on average, which compares with 95 percent in the UK, around 80 percent in the US’ PJM and New Zealand systems, and 73 percent in New England. Moreover, trading volumes can be substantial—up to 4-5 times physical demand (Hedge Market Development Steering Group [HMDSG], 2005). But Anderson et al. (2006) report that most NEM contracts are typically for less than four years duration, mirroring the experience in the US and New Zealand where contract duration is usually no more than three years (Chao et al., 2005; Hansen, 2004). Similarly, low contract durations are reported for the UK, where most trades are for one year only (HMDSG, 2005). Moreover, contract volumes can also be small, with trades in New Zealand representing around just 25 percent of total demand (Electricity Commission, 2006).

**Vertical Integration.** Even where contract markets have developed to a significant degree, and more so where they have not, vertical integration has increasingly arisen. For example, in New Zealand, generation is dominated by five integrated companies (“gen-tailers”), accounting for around 91 percent of generation capacity and 97 percent of total demand (Hogan & Meade, 2007). Similarly, in Spain, four integrated firms account for 93 percent of generation and 97 percent of retail sales (Kuhn & Machado, 2004). And in the PJM system in the US, the six largest retailers account for roughly 70 percent of the retail market and 90 percent of generation capacity (Bushnell, Mansur & Saravia, 2008).

Although such a high degree of vertical integration has been characteristic of, for example, the PJM and New Zealand markets since restructuring in the 1990s, other markets began with a fully disaggregated sector (post-restructuring) that has subsequently moved towards greater integration. For instance, the UK market had three major generation companies and 22 retailers in 1990, but now features substantial vertical integration. In 2005, six major retailers, collectively owning approximately 50 percent of the country’s generation assets, supplied 99 percent of electricity customers (International Energy Agency, 2005). Similarly, Australia began with complete separation of generation and retailing in the mid-1990s, but now vertical integration appears to be becoming a “dominant strategy” both in Victoria (where the first retailer purchases of generation assets occurred) and throughout the National Electricity Market (NEM) (Simshauser, 2008). As of 2007, four major retail businesses, with a combined 75 percent of market share, had ownership stakes in approximately 73 percent of generation assets (NERA, 2007).

In both the UK and Australia, mergers and acquisitions between retailers and generators were not explicitly prohibited by new regulations, and have generally been approved by competition authorities (although initially the Loy Yang merger in Australia was opposed on competition grounds—Gans & Wolak, 2008). In contrast, in markets such as California, regulation brought in during reforms required complete separation, excluding the possibility of any vertical integration (or long-term contracts).

In New Zealand, vertical integration quickly arose as an unintended consequence of simultaneous reforms in 1999 (Evans & Meade, 2005; Hansen, 2004). At a time when the existing contracts market was limited due to the nature of contract requirements imposed on then dominant generator, the government horizontally unbundled that generator, lifted former restrictions on integration between generation and retailing, and required the ownership separation of retailing and distribution. The
newly-created generators rapidly integrated into retailing. The dominance of this model was cemented in 2001, when a winter supply crisis resulted in soaring wholesale electricity prices, placing the largest non-integrated retailer under such a severe price squeeze that it urgently sold its retail base to generators.

2.3. Implications

The incidence of low or decreasing levels of contracting, and high or increasing levels of vertical integration, has led to concerns about market performance. In particular, one fear is that vertical integration enables firms to exercise market power, resulting in higher retail prices (Micola, Ruperez, Banal-Estanol, & Bunn, 2008, Gans, 2007). Similarly, the European Commission (2007) observed that vertical integration appears to reduce liquidity in European electricity markets, and may constrain entry.

However, empirical investigations of market performance have not borne out these concerns. For example, Bushnell et al. (2007) proposed that foreclosure is seldom a problem in practice due to open access and non-discriminatory pricing requirements for distribution and transmission networks. By running simulations based on the performance in the three largest and oldest US electricity markets—California, PJM and New England—they found that vertical integration resulted in the exercise of less market power and contribute to lower retail prices than in an unbundled sector, controlling for horizontal market structure. In contrast, Mansur (2007) found that restructuring in the PJM market did lead to an increase in anti-competitive behaviour by vertically-integrated firms that were large net-sellers, but that vertical integration overall had a mitigating effect on market power.

Conversely, de-integration (unbundling) may be harmful in sectors with existing vertical integration: if the PJM and New England markets had been forced to fully unbundle (as happened in California) retail prices in those areas would have been significantly higher, as would have production inefficiencies (Bushnell et al., 2008). In a similar vein, Michaels (2006), and Sioshansi & Oren (2007), note that forced divestitures may increase incentives for firms to exercise market power. In a review of empirical studies of gasoline refining and sales, Lafontaine & Slade (2007) find that “divorcement” can lead to cost and price increases and a reduction in service quality. Thus, vertical integration may in fact be better for consumers—a conclusion supported by Cooper et al. (2005) in their review of empirical evidence from a range of industries.

Finally, vertical integration may also have a positive impact on investment. As Newbery (2002) explains, competing retailers in a vertically separated market tend to prefer short-term over long-term contracts to mitigate their risk of falling wholesale prices. This increases the wholesale price risks to generators, leading to a decrease in investment. In such a situation, generators will have incentives to vertically integrate with retailers (and horizontally integrate with other generators) in a bid to increase prices to more profitable levels. Thus, vertical integration allows both retailers and generators to manage risks (Chao et al., 2005; Meade, 2001), and in doing so also enables increased investment.

Against this background we now turn to the analytical framework used in later sections to assess the relative merits of contracting and vertical integration in liberalised electricity systems.
3. Analytical framework

In this section we present a framework for reviewing the roles of long-term contracts and vertical integration. We largely draw from the transaction costs theory of governance structure, under which a firm’s choice of governance mechanism (such as contracting or vertically integrating) is based on the costs that firm faces in attempting to transact in the market (Coase, 1937; Williamson, 1985). We also discuss the importance of ownership costs (following Hansmann, 1996) which interact with transaction costs to influence governance choice. This leads in to our more detailed exploration of electricity contracts and vertical integration in Sections 4 and 5.

3.1. A spectrum of governance choices

Sector transactions may be governed under a spectrum of different (explicit or implicit) contracting mechanisms, ranging from markets to firms to bureaucracies (Williamson, 1985). Because each mechanism provides a different level of incentives and involves different costs, each will be “ideal” under a different set of conditions (related to transaction costs, organisational costs, incentives, and ownership). However, experience and theory suggest that a given sector will seldom favour a single ideal governance mechanism; rather, different sector needs and characteristics give rise to complementary combinations of governance mechanisms.

Moving along the governance spectrum: from markets to bureaucracies. Williamson (1985, p. 87) starts from the premise that “in the beginning there were markets” and hence that “Only as market-mediated contracts break down are the transactions in question removed from markets and organised internally.” This suggests a hierarchy among the alternative governance forms (see Fig. 1), with markets being presumed the most desirable form in situations in which they are feasible, then private internal organisation (firms), then bureaucracies (i.e. “political markets”).

Fig. 1: Different governance mechanisms are feasible in different ranges of conditions
However, as Fig. 1 indicates, markets are feasible in a narrower range of conditions than are firms and bureaucracies. As the nature of the transaction changes, so too does the ideal form of governance, leading to a shift "down" the hierarchy (e.g. from spot markets to long term contracts) or re-balancing of the mix of governance mechanisms used (e.g. some spot contracts may remain, but the majority of trading may be through long-term contracts). Which governance mechanism is ideal for a given set of transactions depends on two key factors: (i) the size of transaction costs relative to internal organisation costs, and (ii) the desirability of strong incentives, which incentives weaken as transactions shift away from markets towards bureaucracies.

**From spot markets to longer-term contracts.** In situations with low transaction costs, symmetric information, and limited market power or risk of opportunistic behaviour, spot markets are both a feasible and desirable mechanism for governing exchange – they provide strong incentives and have low organisational costs. As transaction costs increase, and the desirability of strong incentives decreases, other forms of governance become more feasible and desirable. This may involve a shift from spot type contracts to longer term contracts. For example, Adler et al. (1998) note different contract types (such as fixed-fee or cost-plus) are preferred in different transaction cost conditions. These conditions include contract knowledge (“the contact hours the seller and buyer require to learn each other’s requirements to complete the transaction”) and contract impediments (“the inability of the buyer to adequately state contractual terms to the seller, ultimately lengthening contract duration, and changing the technical design”). Fixed-price, spot-type contracts are only feasible when contract impediments are low; as it becomes more difficult to state performance requirements in full, longer-term and more incentivised contracts play a greater role.

**From contracts to firms.** As transaction costs increase further, the shift may be from contracts to firms. For example, in repeated market-based transactions either repeated contracting (such as through the repeated use of spot markets) or long-term contracts are required (Coase, 1937). Such contracts need to be specific about the rights and obligations of the contracting parties in all relevant states of nature. Achieving such specificity – or contractual completeness – becomes increasingly difficult as asymmetric information, uncertainty, bounded rationality, and transaction costs increase. In such situations, internalising transactions with a firm may be a more desirable form of governance. When contracts are internalised they can be less prescriptive and comprehensive. Furthermore, instead of relying on external contract monitoring and enforcement mechanisms, the firm can devise internal mechanisms and tailor these to the particular nature of the transactions.

Authors such as Williamson (1985) extend Coase’s theory further by highlighting the importance of asset specificity in motivating internal organisation (through firms or vertical integration) over market-based transacting. When parties to a contract make long-lived relationship-specific investments, they are then exposed to the risk of ex post opportunistic behaviour in which either or both of them seek to extract rents from the other by renegotiating or reneging on (i.e. "holding-up") the contract once the relationship-specific investments have been made (i.e. sunk). This exposure arises because these investments cannot be costlessly redeployed in the event of hold-up. Such behaviour, if anticipated by the contracting parties, results in less contracting and investment than the parties might otherwise jointly prefer, unless they can otherwise credibly commit to each other not to act opportunistically. Where such commitment mechanisms are too costly to institute, ownership of one party by
the other can be a viable alternative – whence the rationale for either backwards or upstream vertical integration (one party owning its supplier) or forwards or downstream integration (one party owning its customers).

**From firms to bureaucracies.** Finally, at the far end of the spectrum from markets, where transaction costs are particularly high, and high-powered incentives would exacerbate problems such as market power, asymmetric information and opportunism, bureaucracy can be the optimal governance form – it provides weaker incentives and, although it has higher internal organisation costs, these would be offset by the benefits of avoided transaction costs.

**Multiple forms of governance.** Importantly, firms may not choose a single “optimal” governance form based on their overall transaction costs, but rather use multiple forms to address different market needs. For example, Wolak (1996) notes that electric utilities in the United States use both spot and long-term contracts, simultaneously, to purchase the same product (coal, although the situation is similar for gas and oil purchases, and also occurs in the water sector with bulk water purchases). Wolak concludes that, although long-term contracts are good for ensuring supply stability for the majority of demand, “for the most part, plants use spot market transactions to satisfy residual demands due to unforeseen events and are willing to pay even higher prices for and purchase from even more distant suppliers the larger is this residual demand” (p. 164). Thus, spot contracts play an important role in complementing long-term contracts; a role that may also be played in relation to vertical integration, which shares many structural similarities with long-term contracts.

### 3.2. Choosing between contracts and ownership / integration

The particular governance “boundary” that we are interested in for this paper is that between market-based contracting and vertical integration of otherwise contracting partners within a firm. Hansmann (1996) provides a model for considering how market participants decide on which side of the boundary to fall. He acknowledges that long-term contracts have their virtues – in particular, they can be an effective remedy to many of the problems related to reliance on spot transactions. Long-term contracts may be designed to avoid hold-up problems (such as when an investor wants to avoid sinking costs in a long-lived asset only to be held up by its counter-party upon the renewal of a short-term supply contract). Long-term contracts can also be used to allocate specific risks – for example, determining to what extent a supplier is liable for interruptions to supply. Finally, long-term contracts may be used to mitigate adverse selection risks – for example, requiring performance warranties where product or service quality becomes apparent over time.

However, Hansmann (1996) also notes that long-term contracts have costs which, if high enough, will favour vertical integration. He points out that severe problems of contractual imperfection or incompleteness can arise from a combination of asset specificity (i.e. physical and human capital with large sunk costs), temporal specificity (i.e. the requirement that certain products and services be provided at specific times), and high contracting costs. These combined factors explain, for example, why farmer-ownership of dairy processors (i.e. downstream vertical integration) often dominates over investor-owned processors. Milk’s perishability exposes farmers to daily hold-up problems, and homogeneity of interest lowers costs of collective decision making. Similarly, customer-owned rural electricity distribution cooperatives, and upstream generation and transmission cooperatives in the US –
both examples of upstream vertical integration – can be explained in terms of low collective decision making costs (homogeneity of interests) as well as risk of market power abuse by investor-owned firms, or the absence of such firms where investment is not profitable.

Findings in other sectors confirm this preference for vertical integration when contracting costs are high and contracts are necessarily incomplete. For example, from a review of over 100 empirical studies across a diverse range of sectors, Lafontaine and Slade (2007) find substantial evidence that the following factors significantly increase the presence of backwards vertical integration (that is, influence firms to “make” rather than to “buy”):

- Greater specificity of physical capital and of human capital;
- More dedicated and more complex assets;
- Greater site specificity (in other words, when co-location is more important);
- Greater temporal specificity; and
- Greater uncertainty about demand.

Similarly, Lajili et al. (2007) conclude from their review of over 50 studies that “empirical findings generally corroborate the importance of various forms of relationship-specific investments for explaining and predicting vertical integration” (p. 15). In particular, their review confirms the significance of transaction frequency, numbers of trading partners, asset specificity, certainty of asset life span, and demand uncertainty in the decision to vertically integrate. They note that several of these factors are interrelated: for example, higher levels of asset specificity limit the number of trading partners, making it more likely that the firm will be inclined to vertically integrate. Interestingly, Klein & Murphy (1997) propose that vertical integration is more likely in conditions of uncertainty not because of asset specificity-related transaction costs, but because uncertainty makes it more likely that an explicitly specified performance contract “will move outside the self-enforcing range” (p. 420). Similarly, Arrow (1975) develops a model which supports the endogenous evolution of imperfectly competitive vertical integration, even starting with competitive initial conditions. In his model it is uncertainty in the supply of an upstream good – i.e. the risk of a quantity squeeze – that induces downstream firms to integrate upstream.

3.3. Ownership costs combine with transaction costs to affect governance choices

Hansmann (1996) extends the discussion of transaction costs in vertical integration decisions to include a discussion of property rights or ownership costs. Under his scheme, ownership optimally falls to the class of firm patrons (i.e. suppliers,
customers, workers, capital providers) enjoying the lowest combined costs of market contracting and ownership (i.e. internal organisation). For example, if market contracting costs should happen to be highest for a patron class that also has the lowest costs of ownership, then clearly they are the best patrons to own the organisation.

Market contracting costs include the costs of hold-up risks, market power, asymmetric information and strategic bargaining (where a better informed party to the contract negotiates better terms at the other’s expense), problems in credibly signalling patron preferences and the costs of long-term contracting. To this can be added regulatory risk (Helm, 1994; Meade, 2005). Ownership costs, by contrast, include the costs of internal governance mechanisms (i.e. the classic agency costs of Jensen and Meckling, 1976; Fama and Jensen, 1983), the costs of collective decision making, and the costs of risk bearing (which relates to patrons’ diversification and access to capital).

Under Hansmann’s scheme, vertical integration is a mechanism for internalising a range of risks and costs to the firm. By dispensing with the contracting requirements of market-based transactions there are savings in terms of both direct contracting costs and the costs arising from imperfect contracting. Incentives for opportunism are mitigated (though replaced to some extent by agency costs of internal governance), since the costs of that opportunism are borne within the same organisation. Exposure to volatile market prices is removed or reduced, since price changes that disadvantage one part of the firm advantage another part. Information asymmetries give rise to agency costs of internal governance, but once again the costs of asymmetric information and strategic bargaining are internalised to the firm and hence should be lower overall than when they are borne by only one party to a contract.

Although opportunities for double marginalisation may arise through integration, exercise of market power may now become moot, since monopoly pricing in one part of the firm directly increases costs in another part, leaving total firm profit unchanged (Hogan & Meade, 2007). By removing market variables from the firm’s objective function (or simply from view), vertical integration also reduces the risk of (and scope for) adverse regulatory interventions such as price regulation. Integration can also remove the need for otherwise beneficial regulation, since the integrated firm internalises the costs that such regulation would seek to mitigate. It is only when the costs of ownership exceed these combined cost savings that market contracting should remain preferred to vertical integration.

4. Contracts in decentralised electricity systems

In this section we explain and elaborate on the problems that occur in contracts-based electricity markets, identifying why the “conventional” view of appropriate sector structure may fail to produce desired sector outcomes in practice.¹ Reasons for this failure include hold-up risks, adverse selection risks due to market power and asymmetric information, contract market illiquidity (due to features inherent in both electricity systems and electricity markets), and mismatches between generators and retailers in terms of preferred load profiles and relative risk aversion.

¹ A more in-depth discussion of these issues can be found in Meade and O’Connor (2009).
4.1. The need for risk management in electricity markets

Generators face both price and quantity risks in their output and input markets. While risks regarding the price or quantity of capital inputs are amenable to risk management in most developed capital markets (financial crises aside), the same is not always true for other inputs. Supply or price insecurity for gas, such as that in the UK with declining North Sea reserves or other European countries reliant on gas from Russia, create one set of risks. Similar risks arise for hydro generators exposed to variable inflows, storage constraints, and possible environmental regulation affecting off-take (as is the case in New Zealand—Evans and Meade, 2005). Uncertainty about plant availability creates additional input risk. On the output side, generators face highly variable real time demand, as well as daily and seasonal load variations that may not align well with capacity. Real time demand variability can be especially costly where plant is expensive or slow to “ramp up”, such as for coal generation (Wolak, 1996). Moreover, given highly price-inelastic demand in the short-term, small changes in plant availability or demand shifts (e.g. due to weather changes) can result in highly volatile prices. These features can be exacerbated in systems with transmission constraints and locational pricing (such as PJM and New Zealand), since transmission system operators can “constrain off” or “constrain on” generation to maintain system stability, with significant spatial price separation as a result.

Similarly, retailers face price and quantity risks in wholesale and retail markets. They too face input quantity risk, especially where generators are input constrained (such as in hydro systems like New Zealand’s with limited reserves) or have plant prone to outages (e.g. thermal). Such risks also correlate with price risks—often demand is highest in electricity systems precisely when prices are also highest (Chao et al., 2005, Meade, 2003), resulting in highly convex costs. Such convexity is shown to be an important rationale for hedging (Mackay & Moeller, 2007). On their output side, retailers face relatively less price risk, in that smaller customers tend often prefer fixed price contracts. While this exposes retailers to price squeezes if they do not have hedged supplies, in principle they can reduce this risk with appropriate contracts. Like generators, however, retailers remain exposed to real time demand variability, and hence can be exposed to uncertain wholesale spot prices if their contract positions are inadequate.

Large customers, by contrast, face price and quantity risk in their output markets that is less correlated with that of retailers. Moreover, except for aluminium smelters, steel mills and other firms for which electricity costs are a significant part of their overall cost structure, large customers are not exposed to input price risks to the same degree as retailers (for whom energy costs are by far the biggest cost). They remain exposed to some degree, however, to price and quantity volatility in wholesale electricity markets, which affects their need and preference for hedging arrangements. Even when electricity costs are a small part of their overall costs, a secure electricity supply can be essential for their business viability, and hence hedging quantity risks (i.e. securing supply) may matter more to them than hedging price risks.

Contracting can help to manage these risks (see Figure 2). In terms of hedging short-term exposure to price or quantity risks, generators and retailers (and to a lesser extent large customers) are in principle natural counter-parties. By entering into hedge contracts they can lock in some part of their desired demand or assured supply, and at prices that need not be tied to volatile wholesale market prices.
Retailers with spiking demand and fixed price customers will not wish to be exposed to spot market prices for top-ups; whereas generators will not wish to over-commit supply and be forced to purchase at spot prices to make up supply shortfalls. Consequently generators cannot afford to be over-contracted, whereas retailers cannot afford to be under-contracted, possibly creating an overlapping set of contracting preferences. Furthermore there is a natural role for generators that have uncorrelated input or output risks to contract among themselves rather than to bear the risk of having to trade in spot markets to make up supply shortfalls when over-committed, or sell surplus when under-committed. In that way they can hedge themselves against unforeseen plant outages, input shortages or input price hikes. This in turn enables them to issue contracts to retailers and generators over a larger share of their capacity without risking an unhedged exposure.

Large customers and retailers also share some ability to use contracting as a device to constrain generator market power. Requiring generators to commit some part of their capacity via fixed price contracts can reduce their incentive to exercise market power (though see Bonacina, Creti & Manca, 2008, for the range of views), which is one reason why generation privatisations have sometimes involved vesting contracts. Contracting can be used by large customers or organised customer groups, however, as a more proactive tool to address market power concerns, sometimes in lieu of explicit regulation.¹

In terms of supply security it might be thought that generators and retailers (or large customers) will share incentives to contract. Where common security obligations are imposed on all parties this might be so, but common pool resource (i.e. non-excludable but rivalrous) attributes of supply security do not ensure this if free rider problems are severe.²

---

² See Meade (2005a) for a discussion.
4.2. Factors undermining the effectiveness of contracting

Despite the rationale for contracting set out above, contract markets have failed to develop to the extent expected (see discussion in section 1). This is largely because several factors undermine the effectiveness of contracting as a risk-management tool, leading to a reduction in the supply of contracts, increase in wholesale prices, and reduction in investment and supply security.

**Hold-up risks for generators.** The generation investments required for supply security are usually large, sunk and long-lived, creating a preference for generators, their funders and possibly fuel suppliers to secure investment returns for commensurately long periods. By contrast, entry costs in retailing (e.g. to set up energy contracting, billing systems and call centres) are low, as are the costs of smaller customers changing their supplier. This means that retail firms risk losing market share to new entrants, if they fix their input costs by hedge contracts for any sustained period in which spot or contract prices then fall, inducing entry (particularly by "hit and run" retailers). The problem is similar for large customers that contract with generators—if they have competitive output markets (meaning output prices are difficult to increase, despite potentially large increases in input prices), these customers may face hold-up risks that in turn induce them to potentially renege on their contracts with generators.

In such circumstances generators may be unwilling to offer as many long-term contracts as they would otherwise prefer, for fear of hold-up. This reduces the level of investment they can secure against supply contracts, and creates the additional risk of reduced supply security. If such hold-up is sufficiently severe then generators too might find themselves financially at risk, and hence possibly renege on their own upstream contracts (e.g. with fuel suppliers). This is less an issue for generators that are already integrated upstream with fuel supplies (such as coal-fired or hydro stations), but may give rise to cascading hold-up risks with gas or uranium suppliers that have their own large, long-lived and sunk investments. The result may be sub-optimal upstream investments in exploration, extraction or transmission, which would serve to exacerbate the fuel and quantity squeeze risks faced by generators, resulting in reduced investment and supply security, and reduced contracting.

Overall, hold-up risks are most pronounced where the generator has undiversified positions. For example, a generator with geographically dispersed plant and a mixture of fuel types is less exposed to quantity squeezes than a merchant generator with a single plant. Moreover, a generator with a portfolio of supply contracts combining a mixture of loads can smooth out demand volatility that both reduces quantity squeezes and also its exposure to hold-up by any one customer. Facing a smoother load profile means it is also able to write contracts over a greater share of its capacity, all fuel and other supply risks being equal. A merchant generator with a single off-take agreement is clearly in a more vulnerable position.

**Hold-up risks for retailers and large customers.** Retailers and large customers also face hold-up risks from generators—for example when replacing expiring hedge contracts. Retailers and customers with undiversified positions (e.g. a single supply contract) face greater hold-up risks. Retailers with a portfolio of supply contracts for different maturities and/or from different generators diversify their exposure to hold-up by any one generator.
For large industrial customers the hold-up risk posed by generators may be greater than that they pose for generators, particularly if the customer operates in a competitive output market. This is because any given industrial customer may account for a fraction of a generator’s output, whereas any given generator may account for a large share of their supply. However, if a secure electricity supply at a competitive price is important for the ongoing competitiveness of industrial customers (e.g. aluminium smelters) then the location of the customer’s plant would presumably have been decided on the strength of the long-term electricity supply contracts they could negotiate at the outset, which reduces their exposure to generator hold-up.

For incumbent retailers, other sources of hold-up risk may be significant—including the sector reform (liberalization) process. In particular, reforms may involve an unanticipated shift from regulated and smooth pricing (amenable to political lobbying) to volatile or increased prices set by market processes (less amenable to political control). Indeed, Anderson et al. (2006) report that participants in the NEM regard regulatory risk and ongoing reforms as their single greatest risk. In contrast, entrant generators have the benefit of determining whether reformed market conditions enable economic entry.

**Adverse selection risks.** Generator market power (which is greater when contract markets are thin—Green, 2006) and information asymmetries can conspire to create adverse selection problems for retailing firms and large customers. As a result, retailers/ customers are less willing to lock into long-term contracts, particularly if contract prices are temporarily high (e.g. as in the lead-up to peak demand periods in New Zealand when hydro storage is low). Aside from the general difficulty in predicting supply and demand conditions more than a few years out (Anderson et al., 2006)—an example of bounded rationality that hinders effective contracting—these parties can be reluctant to contract with generators whom they believe to possess market power, as well as better knowledge than they about impending outages or fuel insecurity. Where contract prices are perceived to be neither competitive nor reliable then this only increases the risk that they enter into a contract price and then face being out-competed in fixed-price retail markets by competing retailers (or in product markets by competing industrials) who can secure more favourable prices through either superior information or greater negotiating leverage.

**Market illiquidity.** The adverse selection problem, leading to generator hold-up risks, is only made worse by the general problem of illiquid forward markets for which standard arbitrage pricing conditions do not hold. Since electricity is not economically storable, contracts for forward delivery of different maturities are effectively contracts for distinct commodities—energy at defined future times. This thins the relevant markets—potentially exacerbating market power issues—and means temporal hedge markets do not satisfy arbitrage relations based on storage costs. Arbitrage is made even more difficult in systems with transmission capacity auctions (such as system residue auctions for interconnector congestion rents in the NEM, interconnector capacity auctions in the EU, or FTR auctions), due to asynchronous energy and transmission auctions (Anderson et al., 2006). Additionally, the delivery of electricity has a spatial dimension in systems with transmission constraints and either zonal or nodal pricing, which further thins the relevant forward markets.
Mismatch of supply and load. Generator supply and retailer / customer demand may be aligned, enabling straightforward contracting (for example, aluminium smelters may contract with base load generators, or retailers may hedge peak demands by contracting with peak generators). However, any divergence of preferences limits the standardisation of hedge contracts and thus reduces the liquidity of hedge markets. For instance, generators exposed to significant input risk will prefer force majeure or other availability clauses that do not commit them to supply during outages or fuel shortages. Writing contracts that do not align with their supply profiles leaves generators with uncommitted supply that may be harder to contract; thus, generators will seek a premium to write contracts that badly fit their supply profile. In contrast, retailers or industrial customers will generally prefer contracts that follow their own load profile and demand swings, rather than generator supply availabilities. They will also prefer asymmetric instruments, such as options, which limit both their quantity and price risks. Both sides to a contract will therefore need to weigh whether their interests are better served by contracting at a price, or to only contract to a more limited extent so that residual contracting options are preserved (especially when demand or supply conditions are particularly volatile).

Relative risk aversion. Misalignment of contracting preferences is exacerbated by differences in relative risk aversion between the parties. For example, Anderson et al. (2006) report that state-owned generators in the NEM are more risk averse than private generators. This may bias them in favour of greater contracting relative to private generators, but also increases their preference for contracts that do not commit them to supply during outages or fuel shortages. Private generators by contrast may be less inclined to issue contracts (all other things being equal), but be more prepared to offer terms favouring customers who do not wish to bear outage risk. Depending on the fuel type and variety of private and state-owned generators, such differences in risk preferences could create distinct sub-markets for contracts.

Contracting problems become more severe if generators have market power and are also less risk averse than retailers and large customers (because of soft budget constraints for state-owned generators, or because their input and output risks are better off-set than those faced by their customers). In this case generators have less incentive to write contracts for two reasons: they do not wish to reduce their ability to increase prices, and they have less need to hedge in the first place. This does not mean that they will offer no contracts, but rather that contract prices will be higher, contract volumes will be lower and contract terms will be less favourable than if generators did not enjoy market power or have relatively low risk aversion.

4.3. Effect of initial conditions

In addition to the factors described above, the success of contracting as a governance device for managing risks in liberalising electricity systems may also be limited by the initial conditions at the commencement of reforms. Where decentralised systems emerge against a backdrop of significant excess capacity, for example, generators will be more likely to offer contracts for sale and to do so on more favourable terms (e.g. with less protective force majeure clauses or better demand load matching) than if this were not the case. In the same circumstances, however, retailers or large customers face less imperative to sign such contracts, and indeed may prefer not to do given that prices will already be relatively low. As demand grows and excess capacity diminishes, the supply of contracts may shrink at the same time that customers increasingly prefer to hedge against future price rises, unless large customers (for example) are prepared to contract forward to
secure investments offering them greater security of supply. Such forward contracting is facilitated where those customers enjoy some degree of market power in their output markets, or their competitors are also locked into the contracts (e.g. via joint venture).

The extent of pre-existing vertical integration or long-term contracts also plays a role. While vesting contracts are often imposed on dominant generators as a means to constrain their market power as liberalisation unfolds, even in horizontally de-integrated systems pre-existing contracts or vertical integration affect the supply of and demand for contracts. Systems with large industrial customers on long-term contracts signed pre-liberalisation often inherit those contracts as states seek to mirror their commitments in a way that does not crystallise or reveal liabilities to taxpayers. A consequence of this, however, is that the evolution of contracts markets will necessarily be constrained by the scope and terms of such contracts. Similarly, vertical integration, with large customers or otherwise, will also affect contract market evolution since the supply of and demand for contracts will be lower, and hence contract markets less liquid, if significant vertical integration already exists. With these particular electricity sector contracting issues in mind, we now turn to how vertical integration might be a preferable form of organisation, and discuss other possible solutions to the shortcomings of contracting.

5. Vertical integration in decentralised electricity systems

In this section we argue that vertical integration, compared to contracting alone, is a more natural and self-sustaining approach to risk-management in decentralised electricity systems. As a consequence it offers additional benefits over contracting in areas such as investment and security of supply, but also for retail entry and competition. Importantly, this does not mean that all forms of contracting should be eschewed: both spot and long-term markets have a role to play in supporting vertical integration. For example, generators with uncertain fuel supplies will be willing to integrate only if they expect their native generation to mostly be sufficient to meet their embedded load. If their generation capacity is insufficient, they must supplement it with purchases from other generators (at least under some fuel supply scenarios), either on spot or longer-term contract markets. Conversely, if generators under-integrate they will seek to off-load some of their excess capacity via spot markets or contracts, the returns from which will have different risk characteristics. Hence, even integrated generators will seek to access spot and longer-term contracts markets to some degree.

5.1. Lower comparative ownership and contracting costs

As outlined in Section 3, when the costs of internal organisation or ownership are outweighed by the costs of market transactions and of lower investment (resulting from inadequate contracting markets), integration should be preferred over contracting. Importantly, in vertically integrated firms both sets of costs are likely to be lower than for separated firms.

Firstly, vertically integrated firms are likely to have lower costs of internal organisation and ownership, including lower agency costs (which arise when ownership and control are separated), costs of collective decision-making, and costs of risk bearing. In particular, costs of risk-bearing are lower for vertically integrated firms because of:
• their often greater diversification relative to standalone retailers and generators, and
• their larger scale and balance sheets (particularly relative to retailers), which enables integrated firms to access greater capital.1

Secondly, vertically-integrated firms face lower costs of market contracting. As summarised in Figure 3, the potential to avoid high contracting costs provide rationale for both upstream and downstream vertical integration. We discuss these avoided costs further below.

Figure 3: Rationales for Vertical Integration in Decentralised Electricity Systems

5.2. Reduced hold-up risk while sustaining retail competition

In addition to its direct cost benefits over contracting, vertical integration has another clear advantage: it fundamentally reduces the risk of hold-up that can plague non-integrated electricity systems reliant on contracting. By thinning contract markets, vertical integration immediately reduces the scope for the type of retail entry that creates hold-up risks: entry by low-cost, undiversified retailers that are highly reliant on liquid contract markets for supply.2 Even if such retailers did enter a market dominated by vertical integration, they would only be able to access a small amount of contract capacity, placing them at a disadvantage in negotiating with large (integrated) generators. Because of the limited capacity available to purchase and on-sell, the non-integrated retailer will only be able to vie for some portion of the customer base of an integrated generator. Furthermore, an integrated generator needs only to ensure it can recover long-run average costs to ensure its survival against entry. Hence even if it matches the entrant’s prices for its at-risk customer classes it may still remain viable, particularly if it relies on more modest debt

---

1 These arguments are explored further in Section 5.2.
2 This type of entry has other disadvantages, in addition to increasing hold-up risk—for example, Joskow & Tirole (2006) show that “too much” retail competition (amongst non-integrated firms) may lead to inefficiencies such as overinvestment in smart metering.
financing rather than highly-leveraged project financing. Moreover, in principle it can cross-subsidise at-risk customer classes using any non-at-risk customers, further ensuring its financial viability in the face of entry. Overall, these imbalances of size, cost-bearing ability and negotiating power work in favour of incumbent generators, ensuring they are less exposed to hold-up and failure should the entrant retailer succeed, and ultimately promoting increased generation investment.

At the same time, thin contract markets encourage a more sustainable type of retail entry: entry by firms that jointly invest in both retailing and generation, or by incumbent generators that extend into retailing. Of these possibilities, downstream integration by generators seems the most probable. Generators have scale and balance sheet advantages that favour their integration into multiple retailers (as opposed to, for example, integration by multiple retailers into generation, which would raise additional internal organisation costs). Furthermore, new entrants attempting to jointly invest in both retailing and generation (or retailers attempting to integrate upstream) are likely to face barriers to generation ownership.

Capital market constraints are one possible barrier, particularly where capital market development or depth are insufficient for new firms or existing retailers to raise the large sums required for upstream investments into diversified (rather than single plant) generators. Another important barrier is human capital requirements—running generation plants requires engineering (designing, building and operating) and project management expertise, and fuel supply contracts, commonly possessed by generators but which may be difficult for new firms to obtain. Conversely, it is relatively simple for incumbent generators to procure the non-scarce marketing, billing and other operational resources required for retail entry. Interestingly, these arguments suggest that one common concern about integration – that it raises retail entry barriers by thinning contract markets and therefore requires entrants to also have generation capacity – may be misplaced. The concern is natural if it is presumed that non-integrated “bottom-up” retail entry supported by contracts is necessary, but the natural ability of integrated generators to incrementally expand generation capacity and then enter into downstream retailing “top-down” must also be considered.

These entry and risk-management advantages apply not only for generators integrating downstream into retail, but also for those integrating upstream into fuel supply. Such integration reduces the threat of hold-up in (e.g.) third party gas exploration, since generators have large sunk investments in thermal generation that could be stranded if contracted gas supplies do not materialise. It also reduces the potential hold-up risk faced by an independent gas explorer, particularly given uncertainties regarding future (e.g. climate change) policies for thermal generation investments. Furthermore, this form of vertical integration may provide the funding security for such upstream investments. Without embedded customers both generators and fuel suppliers would face downstream hold-up risks that potentially impede such investments.

1 In New Zealand, an additional barrier is state ownership of most of the competing generators (Evans and Meade, 2005). Even if retailers had the capital to acquire generation, they would be reliant on contracting or generator break-up to off-load any surplus capacity acquired, but more importantly the bulk of that generation has simply not been for sale due to political opposition to privatisation. In contrast, retail companies were available for purchase as a consequence of legislation unbundling distribution and retailing, meaning generators were able to buy into such retailers.
5.3. Internalisation of contracting costs

In addition to reducing hold-up risks, vertical integration can internalise a range of price, quantity, and strategic bargaining costs and risks within the firm that otherwise must be managed – less well – through contracting. With any form of vertical integration (upstream or downstream), wholesale price risks become largely moot: if prices are too high then one part of the firm gains while the other loses (Hogan & Meade, 2007). Since contracting is reduced to cover only any remaining uncommitted (or over-committed) capacity (as illustrated in Figure 4), wholesale prices play a much reduced role, and hence any volatility in wholesale prices is of less consequence in the firm’s objective function. Indeed, while spot market prices typically allow recovery of only short-run marginal production costs, integrated firms are concerned more with the recovery of long-run average costs. This goes some way towards addressing the “missing money” problem identified in Joskow (2006), and therefore supporting investment and supply security.

Thus, generators integrate into retailing not to avoid spiking wholesale prices (which they would otherwise prefer to face in contract markets), but rather to provide a natural hedge against low prices in the context of long-lived investments. The same rationale applies for upstream integration by generators into fuel supplies, which can remove exposure to volatile input prices or quantities. This risk is becoming evident in European gas supplies, given the increasing market power exerted by Russia. Upstream integration enables generators to improve fuel supply security or to better manage price risks, reducing their risk of price or quantity squeezes and therefore supporting their generation investments. While the internal organisation costs of upstream investments by generators may be large, hold-up risks in fuel supply mean that contracting may be particularly costly, justifying the costs of integration.

Figure 4: Internalising Wholesale Prices via Vertical Integration
Additionally, while vertical integration thins the markets for contracts, it has the positive externality that it reduces the problems of generator market power and strategic bargaining that otherwise would be commonplace in contract markets. Generators face less incentive to exploit any short-term informational or bargaining advantages they may hold when the bulk of their output is already committed to their (downstream) integrated customers. Integration also protects retailers against generator informational advantages regarding capacity availability, and reduces their adverse selection risk otherwise faced in negotiating hedge contracts. By internalising the costs of retail-level hold-up, retailer investments in peaking capacity are made more viable, which in turn provides credible commitments to funders of and fuel suppliers to such generators.

Looking at market power from another angle, generators may conceivably acquire customers to limit their exposure to buyer market power. While this exposure may not naturally arise in respect of retailers, it may do so in respect of industrial customers. In Australia, for example, aluminium smelters can exert buyer power over stranded generation investments by threatening to relocate when renewing long-term contracts.

5.4. Better supply and load matching

Vertical integration can also provide a long-term match between generator and customer preferences in terms of supply-security and load matching. For any given level of retail entry risk (or output market competition for large customers), diversified and integrated firms are less exposed to the hold-up risk arising with non-integrated customers. For instance, retailers fearful of short-term wholesale price spikes or supply insecurity for which durable contracting arrangements are not available can invest in peaking plant. Large customers with unusual or seasonal load profiles (e.g. dairy processors or pulp and paper processors) can invest in co-generation plant with output that correlates with their production patterns and that also affords them greater control over their achieved level of supply security.

The combined effect of marginalising the integrated firm’s exposure to wholesale price risk, and enabling a better matching of capacity and demand characteristics, means that vertical integration more effectively manages wholesale price risks than does contracting. Moreover, it does this on a secure long-term basis, whereas contracting achieves effective risk management—to the extent it does—only for the contract horizon, which in practice is for only a limited period. Beyond that horizon contracting parties are exposed to renegotiation risks not shared by integrated firms. Internalising the limited wholesale price risks to the firm thus provides a more durable hedge.

5.5. Reduced regulation risks

Integration can also reduce the risk of regulatory intervention. Because wholesale prices are internalised to integrated firms, they become a less reliable signal to regulators and policy makers regarding firm conduct and reform success. As noted above, integrated generators with balanced portfolios have less incentive to exercise wholesale market power; and if customers are on fixed price contracts offered by

---

1 The benefits of integration for wholesale prices are tied to the degree of balance between generation and retailing activities, not the degree of generator market power per se (Hogan and Meade, 2007).
integrated generators then they are insulated from wholesale price movements anyway. Furthermore, the impact of any price regulation that is imposed is weaker (relative to the impact on non-integrated firms), because wholesale prices are relevant to integrated firms only at the margin—so it is only at the margin that any regulatory costs or benefits would accrue. An integrated firm’s more important decision variables—such as fuel stocks and average production costs—are not known to regulators, further complicating any attempt to regulate it. As a result, integrated generators are naturally less exposed to regulatory or system operation interventions that serve to depress wholesale electricity prices.\footnote{On the other hand it should be acknowledged that the greater opacity and irrelevance of wholesale prices under vertical integration may induce regulators to pay greater attention to retail prices. Any attempts to regulate such prices, however, would have to be tailored to the inherent volatility in wholesale prices and fuel supplies, and to institutional arrangements for firms to either manage wholesale price volatility (in contract-based systems) or the margin between retail prices and generation costs (in integrated systems). Any failure to do so could result in a California-like crisis and widespread firm failures.}

Existing regulations may also affect incentives for incumbent retailers to integrate upstream. The incentive for upstream investment will be affected by regulatory or system operation decisions such as price caps or reserves management which affect wholesale price peaks. Stand-alone peaking investment relies on transitory scarcity rents in wholesale prices, and hence can be made non-viable by constraints on price spikes. Similarly the rationale and viability of upstream integration by retailers is also reliant on such (avoided) rents. Thus upstream integration by retailers may not be as sustainable as downstream integration by generators.

5.6. Favourability in systems with transmission contraints

The advantage of vertical integration over contracting in terms of reduced hold-up risk can also arise from an unexpected quarter—transmission constraints. In systems such as the UK’s where there are relatively few transmission constraints, national energy prices and “postage stamp” transmission charges are supported. In turn this facilitates retail-only entry—and hence hold-up risk—at a national level. By contrast, in systems with significant transmission constraints and locational energy pricing significant and often unpredictable price separation can arise between pricing nodes. In the absence of tools to mitigate the risks to generators and customers of such price separation—such as financial transmission rights or the NEM’s system residual auctions—such price separation can expose either generators or customers to the risk of significant under-hedging via contracts. Accordingly, in such systems vertical integration by generators with plant geographically dispersed plant above and below constraints can be a more effective option for managing risks and sustaining investment incentives.

5.7. Favourability in hydro-based systems with fuel uncertainty

Another set of electricity systems in which vertical integration has an advantage over contracting, with the benefit of increased investment, are those in which fuel uncertainty can result in sustained electricity price rises. This is particularly the case in systems in which hydro storage is limited or uncertain, which can sustain surges in wholesale electricity prices, sometimes for months (in contrast with demand spikes, grid failures, or generation outages, which cause only transitory surges).
Where such uncertainty is systemic and potentially prolonged, generators will naturally incline towards contracts with force majeure clauses that protect them against contract breach in the event of non-delivery. In precisely the same situations, however, large customers and retailers would prefer to not face such clauses, unless they have significant ability to curtail load. In vertically integrated systems, by contrast, these risks are internalised to generators, and increases in fixed retail prices (and hedge contract prices) following sustained increases in wholesale prices become a means of supporting new generation investments, though perhaps more in peaking plant than base load generation.

5.8. possible threshold level of vertical integration

It is possible that there is a threshold level of vertical integration – specific to the characteristics of each system – above which vertical integration becomes a necessary form of organisation. If contracting is not widespread in a non-integrated system, then as higher levels of vertical integration emerge in response to contracting shortcomings, this “fixes” prices for the integrated firms’ share of wholesale market trades. In turn this means that any inherent volatility in generation (i.e. in fuel supplies or plant availability) finds its expression in a diminishing pool of wholesale trades at spot prices. As a result those spot prices become more volatile, further increasing the rationale for integration (as a tool for managing price risks, absent effective contract-based solutions). While it can be argued this simply gives rise to a coordination problem that might be resolved in favour of contracting by mandating contracting or prohibiting integration, the arguments in this paper suggest this tendency towards integration is in fact the efficient solution.

5.9. Solutions to contracting problems other than vertical integration

With these advantages of vertical integration relative to contracting in mind, we now briefly consider possible non-integration based fixes to the problems of contracting in decentralised electricity systems.

Regulating for contracts. Regulating for contracts is often a principle driver of contracting in decentralised electricity systems. Vesting contracts are commonly employed to limit market power and provide wholesale price certainty in liberalising systems, particularly while generator market power is seen as an important risk. Virtual power plant agreements are sometimes imposed on dominant generators as a means to reduce their market share while deferring horizontal de-integration or privatisation. Similarly, regulations can require non-dominant generators to sell a minimum share of their capacity via contracts. Long-term contracts can be imposed on generators so that liberalising states can offload legacy long-term supply contracts previously entered into with industrial customers. Also, contracts have been imposed in states where liberalisation processes have not performed in a politically sustainable way, giving rise to reactive interventions (e.g. California, Ontario – see Chao et al., 2005, and Tetrault, 2006, respectively).

The fact that regulation may be required in order to sustain contracts markets begs the question as to the market failure such regulation is intended to remedy (see Willems & De Corte (2008) for one such rationale). This is especially so if vertical

---

1 This argument is analogous to the idea in finance that increased financial leverage increases the riskiness of returns to shareholders for any given underlying level of business risk.
integration has been dismantled or prohibited as part of a liberalisation process. If either is the case then a potentially endogenous approach to risk management has been artificially precluded, and so the basis for regulating for contracts is potentially faulty. Where vertical integration is permitted, however, but does not arise, a more natural question is why such integration might be impeded, rather than why contracts need to be encouraged. If there should happen to be natural impediments to vertical integration then there is a clearer rationale for contracting, and hence regulating for contracting might have a more sound rationale.

Reducing retail competitiveness. It is curious that hold-up problems have arisen in electricity sectors as a consequence of retail competition, given that competitive entry into retailing has been less than comprehensive (Defeuilley, 2009). Indeed, aside from barriers to retail entry such as transmission constraints and vertical integration, there are a range of reasons why customers have proved reluctant to switch to new suppliers. Among these are bounded rationality (e.g. customers’ inability to understand new retail contracts), and a lack of customer motivation (e.g. because power bills are often only a fraction of overall household expenditures, so any switching savings are relatively modest). Retail customers can also be reluctant to switch suppliers if only short-term contracts are on offer in case they face price increases after switching (although Defeuilley, 2009 reports that the gains from switching tend to be sustained, with incumbent prices typically exceeding entrant prices over time). Other explanations include the risk of disconnection from changing suppliers, loss of loyalty benefits, or relative unfamiliarity with, and hence possible distrust of, new suppliers. While some “active” customers are highly price-sensitive and inclined to switch (Amundsen & Bergman, 2006), most customers are “inactive” and remain loyal to their incumbent supplier.

Additionally, states experiencing the highest levels of customer switching are often also those implementing measures favouring entry. For example, Defeuilley (2009) reports that Texas, Victoria and South Australia all implemented price regulations designed to encourage entry, with incumbent Texas suppliers required to offer a regulated “price to beat” until their franchise market share fell below 40%. Hence, to the extent that retail entry in such jurisdictions has resulted in hold-up problems, this may be a consequence as much of pro-competition regulation as it is of flaws in contracting or vertical integration.

Such considerations aside, some authors have proposed the retention or reinstatement of retail franchise areas as a solution to hold-up problems created by retail entry (Chao et al., 2005, Roques, 2008, Newbery, 2002, 2005). Doing so means retailers have locked-in customer bases that are insulated from competitive predation. This in turn reduces their incentive to renege on contractual commitments to generators, and affords them greater protection against predation when entering into longer-term supply contracts. Such reduced hold-up risk induces generators to commit a larger share of their output via contracts, which can also reduce their incentive to exercise wholesale market power, thereby further supporting the use of contracts as a risk management device. Furthermore, reduced hold-up risks and greater use of long-term contracts with retail counterparties that are less likely to face financial distress means generators are better able to manage their investment risks, lowering their cost of capital and increasing their access to investment capital. By locking in customers, retailers should also be able to enter into less diversified contracts with generators whose capacity better matches their load requirements (e.g. merchant peaking plant), enabling more tailored contracting.
Such measures may be necessary in electricity systems otherwise lacking natural constraints on retail competition (such as transmission constraints), other causes of customer "stickiness", or sources of scarcity rents supporting investment (e.g. as in hydro-exposed systems). However, they appear to be a blunt instrument to achieve the desired end. For example, similar hold-up risks can also arise in respect of large customers operating in competitive output markets (e.g. aluminium smelters), but prescribing entry barriers in such markets would seem an unnecessary and unnatural response to support generation investment. Additionally, the same considerations discussed in relation to regulating for contracting apply, since retaining or reinstating retail franchise areas are examples of such regulation. This approach shifts the problems of risk management into the regulatory domain, which raises its own hold-up risks and hence threats to investment and supply security. If it succeeds, then the risk is that it does so at the undue expense of consumers.

**Other Demand-Side fixes.** At least some of the problems presented by contracting can be mitigated by demand-side fixes not requiring constraints on retail competition. In particular, there are long-standing reform areas such as increasing real-time demand-side responsiveness that help to smooth load profiles, reduce demand peaks and uncertainty, mitigate the impacts of generator market power, reduce generation investment needs, and enhance supply security. If retail customer demand can be better managed so as to avoid peak demands, this reduces the need for often idle peaking generation capacity (and hence also the possible adverse effects of wholesale price regulation and system operating rules on investment). By inducing greater demand-side responsiveness, such as by affording retail customers with options to profit by voluntarily curtailing load during times of peak demand, the transitory exercise of generator market power during times of tight system conditions results in lower welfare loss (Evans & Meade, 2005). Such responsiveness could be achieved either ex ante via load limiting devices (Doorman, 2003), or in real time via retail power exchanges. By reducing the possible welfare losses from generator market power this also reduces the need for regulatory interventions such as wholesale price caps, and hence supports the lower required levels of capacity investment. In unintegrated systems this should also be reflected in lower retail prices, all other things being equal, thus reducing the need for further retail competition (which would worsen any hold-up risks). Hence a suite of such measures should both reduce contracting costs and improve supply security and investment incentives for any given level of retail competition. Where they are combined with vertical integration this should be more so, and they also should ameliorate concerns that vertical integration creates barriers to retail entry.

Given these considerations we conclude that unless the costs of internal organisation are severe, some form of vertical integration is likely to more naturally dominate in decentralised electricity systems over contracting as a mechanism to manage wholesale price, market power, information asymmetry and investment risks. This leaves complementary roles for spot market and long-term contracting, and de-emphasises hold-up risks from retail entry.

**6. Conclusion. Policy implications**

Decentralised electricity systems have been premised on the efficiency and welfare benefits of competition, and have intentionally shifted investment risks from taxpayers and consumers to investors. Where this coincided with excess capacity in generation, this has supported the evolution of contracts markets and retail entry.
However, as capacity margins shrink post-liberalisation, the shortcomings of reliance on contracting – hold-up risks, risk-management mismatches between generators and customers, and market power, asymmetric information and strategic bargaining risks – have become apparent. These shortcomings hinder new investment and threaten the perpetuation of sub-optimal capacity margins. Absent complementary or alternative devices (such as vertical integration, forced contracting or capacity obligations) to compensate for such shortcomings, this threatens supply security and hence the political sustainability of liberalisation, and risks the introduction of destabilising interventions that worsen those shortcomings, or outright re-centralisation.

Contracts markets have often arisen more by design than evolution, for example reflecting long-term contracts entered into pre-liberalisation by states with industrial customers. Often they are imposed to facilitate a transition to generator competition while generator dominance is still a concern. Moreover, vertical de-integration has often been a concomitant of horizontal de-integration, artificially (if unintentionally) limiting the role of vertical integration as a risk management device. These have all served to give contracting a prominence that it may not naturally deserve, and eventually highlighted the difficulties in sustaining contracting. This raises the question whether contracting is simply a stepping stone to greater use of vertical integration (or other mechanisms not sharing the shortcomings of contracting). Indeed, if contracting is such a stepping stone, then beyond a certain threshold level of vertical integration it is possible that integration will naturally become the dominant mode of organisation in any given electricity sector.

As such limitations are revealed the liberalisation pendulum appears to be swinging away from pure reliance on contracting in favour of mixed approaches involving either endogenous vertical integration or impositions such as capacity mechanisms. Such a swing coincides with an evolution in the political imperatives surrounding liberalisation. While safeguarding consumers against generator market power was commonly a political necessity accompanying decentralisation, ensuring supply security – especially in the face of notable reform failures – is increasingly the political priority. Where reformers have retained confidence in decentralised solutions this has involved a tolerance of greater vertical integration. Otherwise it has involved imposed solutions such as capacity obligations, which contain their own inherent shortcomings (Meade, 2005).

Importantly, this evolution requires re-evaluation of the optimal degree of competition in electricity systems, particularly in retailing. This paper stresses that the pursuit of unfettered retail competition – as well as the elimination of other aspects of electricity systems limiting competition (e.g. transmission constraints) – exacerbates hold-up risks, impedes generation investment and threatens supply security. Hence textbook pure competition should not be the policy aim. The reinstatement of retail franchise monopolies is an extreme alternative, however, and may only be justified in the absence of other inherent barriers to retail competition, or by a lack of other system characteristics supporting generation investment. Where such barriers (such as “regionalisation” due to grid constraints) arise, or system characteristics (such as prolonged rather than only transitory price rises as in hydro-exposed systems) otherwise support investment, less extreme levels of retail market power may be sufficient to resolve investment and supply security issues. This is particularly so where systems involve significant degrees of vertical integration to complement or substitute for contracting, in which the hold-up and other risks inherent in contracting are less prominent.
Notably, the tradeoffs in electricity systems based around contracting suggest a lower optimum level of retail competition than in those allowing integration. A key challenge for policymakers wishing to stop short of outright reinstatement of retail franchises is to identify the degree of retail market power striking the optimal balance between protecting consumers while maintaining efficient investment incentives. Where those incentives are distorted by interventions such as price caps – particularly in vertically integrated systems where such caps are most likely redundant – the removal of such distortions is an obvious starting point. If there is to be a bias in policymakers’ approach, it is likely to best be in favour of supporting investment incentives at the possible risk of short-term retail market power. Policy parameters can then be incrementally refined if it proves that investment incentives are overly generous.

The difficulty with the alternative approach is that undue emphasis on controlling retail market power can result in inadequate investment in many systems. The objective should be to encourage desirable long-term decision-making even if this involves short-term biases, since the alternative is to avoid those biases while forestalling or otherwise distorting those long-term decisions. Where other solutions to market power can be implemented without raising retail entry risks – such as measures to improve demand side responsiveness – those are worthy of exploration in their own right. They also de-emphasise the importance of retail market competitiveness, and further complement the benefits of vertical integration over contracting in electricity systems.

Abandoning the relentless pursuit of retail market competition involves breaking from one of the important premises of reform – that retail prices will be necessarily reduced or constrained by greater competition over time. Politicians are required to facilitate or allow the emergence of an acceptable level of retail competition, and trust that the decentralised investment signals this produces are sufficient to elicit acceptable generation capacity and supply security. Alternatively, they can substitute for those signals by imposing investment requirements through capacity obligations or otherwise, and in doing so possibly create the need for greater interventions should the signals prove inadequate, particularly given the greater regulatory risks they create. The merit of greater use of vertical integration instead of contracting is that it better insulates electricity systems from their inherent imperfections, as well as from imposed imperfections (e.g. price caps) introduced to address the inherent ones.

Contracting, just like spot markets, continues to play an important role even in vertically integrated systems. Uncertainties in both supply and demand mean integrated generators cannot continuously maintain perfect balance between capacity and embedded load, and hence can integrate only to some degree. Relying simply on spot trading to compensate for any real time imbalances exposes integrated generators to more short-term price volatility than is likely to best enable risk-management over their long-lived generation portfolio. Combining such spot trading with a portfolio of longer-term contracts enables better management of investment and risks for that part of the generator’s portfolio that cannot be perfectly balanced through integration. Integration is naturally the dominant risk-management approach where the costs of contracting are large, or simply greater than the costs of internal organisation. Just as competition through market mechanisms is expected to enable lowest-cost generation to displace higher-cost generation, policymakers should anticipate that market mechanisms will also enable the efficient mix of contracting and integration to emerge where each is permitted.
There is currently a lack of formal theoretical work identifying the optimal degree of retail competition supporting efficient investment levels and endogenously-determined optimal supply security. This paper presents a natural framework for modelling the optimal use of contracting and integration as a question of mechanism design, given different degrees of retail competition. Such formal modelling presents rich possibilities for future work.

References


Arrow, K. J. (1975). "Vertical integration and communication". The Bell Journal of Economics, 6(1), Spring, 173-183


