A Theory of Optimal Contract Duration
with Application to Outsourcing

Matthew Ellman∗
Universitat Pompeu Fabra

October, 2004

Abstract

What does incomplete contract theory tell us about contract duration? This paper analyzes the costs and benefits of contract length in a multiple investment model that captures the problem of side-compatibility: a long-term contract impedes valuable market forces if adaptations cannot be implemented by market trading alongside the contract. The contract then causes holdup of adaptation investments. The main predictions are that contract length should increase with the importance and specificity of self-investments, and decrease with the importance of cross-investments, wasteful investments and adaptation investments with low side-compatibility. I apply the side-compatibility concept to resolve three empirical puzzles in outsourcing.

JEL CLASSIFICATION NUMBERS: D23.

KEYWORDS: Contract length; market forces; incomplete contracts; holdup.

∗I thank Erik Beulen, Ramon Campabadal, Oliver Hart, Ashok Kaul, Eric Maskin, Elisenda Monforte, Michael Raith, Jan Sarsanadas, Andrei Shleifer, Jérôme Vandenbussche, Abe Wickelgren and seminar participants at Harvard Kennedy School, Kellogg School of Management, Northwestern, Tilburg University, Universitat Pompeu Fabra and University College London for helpful comments. I gratefully acknowledge financial support from the European Commission (HPMF-CT-1999-00317) and the Spanish Ministry of Science and Technology (BEC 2003-00412). Comments gratefully received at Matthew.Ellman@upf.edu.
1 Introduction

Transactions Cost and Incomplete Contracts theorists have shown how long-term contracts help protect relationship-specific investments (see Williamson, 1975 and 1985, Klein, Crawford and Alchian, 1978, Grout, 1984, and Hart, 1995), but we know little about the costs of extending contract length. This paper focuses on a simple yet powerful idea: long-term contracts may obstruct beneficial market forces. My central conceptual contribution is to show how and when this market-shielding cost arises. I also formalize two other major costs of contract duration. I then represent the costs and benefits of contract duration within a single theoretical model. Solving the resulting tradeoffs allows me to predict optimal contract duration and resolve several empirical puzzles. My framework also throws light on the practical challenge of designing flexibility into long-term contracts.

Contracts vary greatly in length. Multi-year contracts are common in, for example, construction, contract manufacturing, distribution and franchising. Outsourcing in Information Technology (IT) presents an instructive showcase.\(^1\) For instance, IBM Global Services sells IT services to Nokia, Hertz, American Express (Amex) and Deutsche-Bank (DB) under 5, 6, 7 and 10 year contracts, respectively. In particular, I use IBM’s $4bn deal with Amex (see Weiss, 2002) to illustrate the main ideas of the paper. Signed in February of 2002, the contract essentially obliges IBM to manage and maintain Amex’s Information System (IS) at a predetermined price. I call this exchange the “basic trade.”\(^2\) The contract duration is 7 years - it enforces the basic trade from 2002 till 2009 - but through renegotiation and renewal, Amex could have paid IBM to provide the same services under a sequence of longer or shorter contracts. It is therefore natural to ask: what are the costs and benefits of increasing contract length?

I begin with the benefits. IBM invested in organization and planning to lower its costs of managing Amex’s Information System. Many of these investments are partly specific to Amex. For instance, IBM coached 2000 workers who transferred from Amex to IBM in 2002. This reduced IBM’s cost of servicing Amex, but the transferred workers’ specific knowledge of Amex does not help IBM in servicing other clients. So IBM must sell services to Amex if it is to fully exploit its cost-reducing investment. Under short-term contracting, Amex could therefore extract a share of the cost-reduction by refusing to buy unless IBM lowers its price.

\(^1\)The IT outsourcing market was worth over $569bn in 2003, up 6.2% from 2002 (Pruitt, 2004). US government IT outsourcing hit $8.5bn in 2003 and is likely to exceed $15bn by 2008 (Input, 2004). The UK National Health Service alone awarded five IT contracts in January 2004, each 10 years long and worth over $1bn.

\(^2\)The contract actually specifies a range of prices (see section 6). It covers infrastructure (data center, networks and desktops) and application software, but is basic relative to subsequently discovered adaptations.
This “holdup” by Amex reduces IBM’s investment returns, so it causes IBM to underinvest. The long-term contract serves to protect IBM’s investment against holdup by fixing the price Amex pays for the basic trade. Lengthening the contract raises IBM’s incentives by increasing the duration of this protection.

The long-term contract can also protect IBM-specific investments by Amex. For instance, when Amex learns to make better use of the basic IT from IBM, the contract prevents IBM from holding up Amex by raising the trade price. In general, long-term contracts protect specific investments that raise the investor’s own payoff from the basic trade. These are called “self-investments” (see MacLeod and Malcomson, 1993). Motivating self-investments is the key benefit of long-term contracting. I now turn to my contributions which deal with the costs.

My primary contribution is to identify the market-shielding cost of contracting. Long-term contracts introduce a serious problem. They often obstruct beneficial market forces. Consider what happens when Amex invests in market research and discovers a valuable IT innovation. For instance, Amex planned a web-based expense reporting service for its corporate card customers in 2003. To exploit this innovation, Amex needed an adapted IT service with new software and third-party web access. The long-term contract did not oblige IBM to provide the adapted system, since it was not included in the basic trade. So Amex had to adapt its contracts to exploit its investment, which I therefore define as an “adaptation investment.” Amex had two choices. Either negotiate a deal with IBM or pay a competitor to make the changes.

Without the long-term contract, both alternatives would have been reasonable. Amex could have threatened to turn to a market of alternative providers, because its adaptation investment was not fully specific to IBM. An IBM competitor such as EDS could have provided the adapted IT service in IBM’s place. Market forces would then have limited the price of adaptation, giving Amex strong incentives to invest. Unfortunately, the long-term contract shielded the relationship from these market forces. The contract reduced market access because Amex could not credibly threaten to buy the adapted service from EDS alongside the basic service from IBM. The adapted service mostly duplicated the basic service, its additional value did not justify EDS’s avoidable costs of substituting for IBM whose investment costs are sunk. Amex was unable to exploit its investment, because it depended on negotiating a contractual change with IBM. This allowed IBM to hold up Amex and caused underinvestment.

Long-term contracts do not always cause holdup of adaptation investments. Indeed the above interpretation of Amex’s adapted service was intentionally extreme. I define the “side-compatibility” of an adaptation investment with a long-term contract to be the fraction of
investment returns that the investor can credibly exploit by trading alongside this contract. Long-term contracts pose no problem for fully side-compatible investments, because the investor can still threaten to turn to the market. Side-compatibility is largely determined by whether the adapted service can or cannot be separated into the basic service and an adaptation service. In the non-separable case, side-compatibility is often zero, but the greater value of the adapted service may still justify paying the duplicated costs of service provision or the penalties for breaching the long-term contract (as commonly occurs after mergers). In the separable case, side-compatibility is usually higher, but is limited to the extent that a single provider of the basic and adaptation tasks can reap economies of scope. For instance, if Amex can buy the new software service from EDS as an “additional service” (complementing the basic service from IBM), there is no duplication of the basic service costs, but EDS may have to pay the same fixed costs of learning about Amex whether EDS entirely replaces IBM or simply provides the additional service alongside IBM’s basic service. Separate provision can also cause problems of coordination and interference: IBM could refuse to provide user-support or third-party web access for software developed by EDS, or abuse its power as IT host to study EDS’s proprietary code. In each case, side-compatibility is reduced.

Side-compatibility is endogenous to the organization of trade, and the design of contracts. I now explain how the concept of side-compatibility can resolve the three empirical puzzles in the contexts of multi-sourcing, buyer dedication, and exclusive contracting. To understand the growing popularity of multi-sourcing, suppose Amex had split its long-term IT needs between two vendors back in 2002. For instance, Amex could have contracted IBM to maintain only its servers and networks, while contracting EDS to manage its help-desk, applications and desktop environment. This would have wasted possible economies of scope by making both EDS and IBM sink fixed costs in learning about Amex. The advantage is that EDS and IBM would then have competed for more of Amex’s subsequent adaptations. More generally, multi-sourcing raises side-compatibility, because the multiple vendors sink their fixed costs in advance and each can exploit some economies of scope if performing adaptations ex post.

This multi-vendor result resolves an interesting empirical puzzle. Transaction cost theory

---

3E.g., Lacity and Wilcocks (1998) report how Diverse1 migrated to a client-server technology alongside a mainframe service contract.

4Tirole (1990, page 54) defines long-term contracts to have breach penalties that enforce future trade performance, but breach may be stochastic. See subsection 6.2 on optimal breach measures.

5See Lacity and Wilcocks (1998) and the contractual exemptions I describe in section 6.

6An example of this “multi-vendor approach:” Procter & Gamble recently rejected a mega-contract with EDS in favour of smaller contracts with HP, IBM and Jones, Lang & Lasalle. See also section 6 on “selective outsourcing.”
informally predicts shorter contract durations for companies that buy from multiple vendors - see González and López (2002). Their logic is persuasive: the multiplicity of active suppliers suggest high competition, so investment specificity and the resulting need for long-term contracting is low - proposition 5 below formalizes this. However, the evidence from electronics outsourcing (López and Ventura, 2001, and González and López, 2002) suggests that multi-sourcers (buyers with multiple vendors) tend, if anything, to use longer contracts. My theory explains the puzzle: increased side-compatibility allows multi-sourcers to write longer contracts. The side-compatibility effect counteracts the specificity effect.

The concept of side-compatibility also applies to adaptation investments by sellers (vendors). For instance, IBM may develop a more secure IT service (e.g., by investing in asynchronous chip technology - see Economist, 2001). IBM has a huge capacity so it could doubtless earn a market reward alongside the basic trade with Amex, by selling this adapted service to other buyers. However, side-compatibility is limited for a smaller IT company when an intensive long-term contract ties up its production capacity. Limited seller capacity for trade plays the parallel role to limited buyer need for trade.7

Companies usually design their long-term contracts to employ only a fraction of their capacity, but side-compatibility is limited for a seller that dedicates most of its capacity to a single buyer. This can resolve a second empirical puzzle. Kerkvliet and Shogren (2001) measure the dedication of coal companies to satisfying their contracts. These scholars predicted that contract length would correlate positively with dedication since dedication is a standard proxy for relationship-specificity, but they found that the most dedicated coal companies actually tend to write shorter contracts. The side-compatibility concept identifies a countervailing force that can explain their puzzling result: dedication lowers side-compatibility, thereby raising the cost of long-term contracting.

Long-term contracts limit side-compatibility more directly if they use exclusivity clauses.8 For instance, if Amex committed to buy all its IT from IBM, it could never threaten to seek an IT adaptation from a third party. This can explain an empirical puzzle in the franchising literature. Exclusive territories are sometimes used to complement long-term performance contracts in protecting specific investments, so Bercovitz (2000) predicted a positive correlation, but found the opposite. The side-compatibility concept explains the puzzle: contracts should be shorter when exclusivity is needed, because exclusivity lowers side-compatibility.

Long-term contracts can also be costly in non-market settings. First, they reduce incentives

7Economies of scope are again important, e.g., partner-specific fixed costs may tighten capacity constraints.
8IT contracts often try to do the opposite. See e.g., GM’s “right to solicit bids” in its 10-year $32bn contract with EDS, and M&I’s obligation to cooperate with third-parties in its Tri City contract (CORI).
to make investments with positive contractual externalities. For instance, when IBM improves storage efficiency this benefits Amex since their long-term contract has a pay-as-you-use formula (pricing per unit of server space actually used). Unlike IBM’s cost-cutting self-investment that raises its own payoff, this is a “cross-investment” by IBM because it benefits the other party. The long-term contract reduces IBM’s incentive to improve storage efficiency, because it prevents Amex from punishing low efficiency with a termination threat. Second, long-term contracting can encourage wasteful self-investments. For instance, IBM might try to hide low quality aspects of its service or litigate Amex - indeed, EDS litigated Xerox over their 10 year contract. Similarly, long-term contracts can cause over-investment in self-investments that have negative cross-investment effects such as lowering quality. My model shows that short-term contracting avoids all these problems, but at the cost of introducing parallel problems for investments with “cross-general” effects (on other party’s market payoffs).

The secondary contribution of this paper is to formulate a generic model of trade between two parties making multiple investments. My first set of results characterizes how different investments respond to changes in contract length. I classify investments into type 1 investments that rise with contract length and type 2 investments that fall. For instance, self-investments and cross-general investments are type 1, while adaptation investments with limited side-compatibility and cross-investments are type 2. I then use these results to determine how contract length optimally trades off investments based on relative desirability. My second set of results can be summarized as: (1) contract length increases with the desirability of type 1 relative to type 2 investments; (2) contract length rises with the relationship-specificity of type 1 investments and with the side-compatibility of type 2 investments; (3) contract length always exceeds the least gestation period of type 2 investments, but is non-monotonic in the gestation period of type 1 investments.

The rich microeconomics literature on contract design (see Holmström and Milgrom, 1991, or Salanié, 1997) tends to ignore a key factor: discoveries are made over time. I follow the “incomplete contracting approach” (see Hart, 1995) in emphasizing unanticipated discoveries, but my results do not require shifts in residual rights of control. In section 4, I use Hart et al.’s (1997) over-investment insight and Che and Hausch’s (1999) model of how long-term contracts harm cross-investments. My focus on the time dimension of contracting complements the literatures on damage measures (see Rogerson, 1984, and Shavell, 1984) and variable

---

11Contracts exchange obligations as well as rights. I isolate the role of trade obligations imposed by performance contracts, in contrast to Hart, Shleifer and Vishny’s (1997) preceding analysis of service privatisation where long-term contracting is tied to private ownership.
quantity contracts (see Edlin and Reichelstein, 1996). Farrell and Shapiro (1989), MacLeod and Malcombe (1993) and Guriev and Kvassov (2004) who also study contracting on trade in multiple periods are complementary. None of these papers capture the market-shielding problem because they do not allow for adaptation investments.\textsuperscript{12}

The transaction cost literature (see Williamson, 1985) presents vital insights and empirics. Joskow’s (1987) empirical support for the idea that contract length should increase with relationship-specificity (as proxied by limited competitive alternatives) is now a classic, and has been confirmed in several industries (see Masten and Saussier, 2002). My propositions 4 and 5 formalize this idea. Furthermore, because uncertainty and complexity raise the importance of ongoing adaptations (and raise the risk that contracts motivate undesirable investments), my theory explains the growing evidence that these factors lead to shorter contracts.\textsuperscript{13} Finally, my analysis refines the important idea that long-term contracts cause inflexibility. In my model, long-term contracts reduce competition for adaptations with limited side-compatibility. This raises the cost of implementing adaptations and (since market forces generate information improving negotiation efficiency) raises the risk of failing to adapt over time.\textsuperscript{14} However, long-term contracts only cause adaptation costs to the extent that side-compatibility is limited, and this refinement is critical for explaining the three empirical puzzles noted above.

The paper is organized as follows. Section 2 presents the basic model. Section 3 solves it for optimal contract length. Section 4 introduces cross effects and wasteful investments. Section 5 extends the number of trading periods. Section 6 endogenizes side-compatibility through empirically motivated features of contract design, and applies the theoretical results to explain the puzzles from prior empirical investigations. Section 7 concludes.

## 2 Basic Model

This section introduces a simple model to analyze how imperfect contracting affects efficiency in a bilateral relationship embedded in a market. I denote the two central actors, P for principal and A for agent, though buyer-seller interpretations are also valid. P and A trade goods and

\textsuperscript{12}MacLeod and Malcombe (1993) do include a general investment, but it is also protected by the long-term contract so there is no market-shielding problem. Guriev and Kvassov’s (2004) contemporaneous and complementary analysis treats the over-investment and cross-investment problems. They only treat a single investment but they allow for ongoing investment. They derive dynamic results when exogenous renegotiation costs cause inflexibility, whereas I endogenize how contracts cause inflexibility through market-shielding.


\textsuperscript{14}For instance, 7-Eleven and Nokia waited until their IT contracts expired before adapting their IT.
services over time. In this section, I present a simplified timing: P and A can negotiate an initial contract when they meet in stage 0 (2002 in Amex and IBM’s case). In stage 1, they invest to raise trade surplus. In stage 2, after observing their trade values, they finalize the contracts that will govern trading in stage 3.

**Contracting.** Advance contracting is restricted by P and A’s bounded ability to think up ways to enforce future trading: In stage 0, P and A can only choose between writing a null contract, \( \Phi \), and a single performance contract, \( X \), inducing non-trivial joint trade in stage 3. Over stage 1, they learn better ways to trade or contract over trade. So in stage 2, they can choose an “adapted” contract, \( Z \) (generating the highest surplus), as well as the “basic” contract \( X \) and the null contract, \( \Phi \). Contracting is *long-term* when P and A agree on a performance contract \( (X) \) at stage 0, and is *short-term* when they select the null contract \( \Phi \), leaving trade negotiation to stage 2. In the multi-period extension (section 5), I extend by defining contract length as the amount of time during which the contract enforces performance.\(^{15}\)

**Default payoffs.** P and A always have symmetric information and I assume that stage 3 renegotiation leads to a fixed and equal (except in section 4) split of any negotiation surplus over the default outcome where P and A do not adjust their contract agreed in stage 0.\(^{16}\) Under short-term contracting, P and A’s default outcome is trade with alternative partners: their joint contract \( \Phi \) leads to \( (\phi \text{ denoting absence of joint trade, and} \phi' \text{ denoting P and A’s optimal market trades.} \) Under long-term contracting, the default outcomes are \( x \) and \( x' \) where \( x \) denotes the basic joint trade induced by \( X \), and \( x' \) denotes the pair of market trades chosen by P and A in response to \( X \).

The default payoffs also depend on P and A’s investments. A typical investment, \( e_j \in \mathbb{R}_+ \) by \( j \in \{ P, A \} \), increases P and A’s optimized stage 3 trade surplus by \( W_j (e_j) \), but it only raises \( j \)’s default payoff (from \( \phi' \)) under \( \Phi \) by \( \gamma_{ej} W_j (e_j) \), and it only raises \( j \)’s default payoff (from \( x \) and \( x' \)) under \( X \) by \( \sigma_{ej} W_j (e_j) \) (from \( x \)) plus \( \kappa_{ej} W_j (e_j) \) (from \( x' \)).\(^{17}\) \( \gamma_{ej} \) represents the investment’s *generality*: \( \gamma_{ej} = 0 \) for a fully specific investment as this has no value in market trading, \( \phi' \). \( \sigma_{ej} \) represents the self-investment effect: \( j \) benefits from its investment through the basic trade, \( x \), induced by long-term contracting on \( X \).\(^{18}\) This measures the investment’s

\(^{15}\)Here, any performance contract spanning the investment period is called “long-term.” The optimal *duration* of the relationship is exogenously fixed (as the interval from stage 0 to 3) under the simplifying assumption that \( Z \) always involves joint trade.

\(^{16}\)Common extensive-form bargaining games give this Nash-Bargaining solution.

\(^{17}\)I assume linearity and separability to simplify the exposition.

\(^{18}\)This occurs when the investor reduces its cost of satisfying \( X \) (e.g., IBM coaching) or raises its benefits from its partner satisfying \( X \) (e.g., Amex learning to coordinate with IBM).
direct-compatibility with the contract. $\kappa_{ej}$ represents the side-compatibility effect: $j$ benefits from the investment by trading $x'$ alongside the contract $X$. I define the sum, $\psi_{ej} = \sigma_{ej} + \kappa_{ej}$, to capture overall compatibility with $X$ (contract-compatibility). The contract length tradeoff depends only on $\psi$ and $\gamma$, but I decompose $\psi$ to analyze how contracts shield out market forces and lower incentives.

Market-shielding occurs when $\kappa < \gamma$. I define $e_j$ to be a pure adaptation investment if it has $\sigma = 0$. In other words, it needs an adapted contract\(^{19}\) so it gives no return under $X$.\(^{20}\) An adaptation investment can still be compatible with contract, $X$, if $\kappa$ from side-trading ($x'$) is positive. However, for adaptation investments, $\psi(= \kappa)$ is usually less than $\gamma$, because side-compatibility is lower than generality whenever the obligation to trade $x$ restricts market access.\(^{21}\) The conditions for this market-shielding can be distilled from the introductory examples: (1) $\kappa < \gamma$ for investments requiring an adapted trade that substitutes for $x$, because this implies duplication waste; (2) When the adapted trade can be separated into the basic trade $x$ plus an adaptation, $\kappa$ is higher but still less than $\gamma$ to the extent that separating these tasks (between the alternative and original providers) wastes economies of scope, harms coordination and permits interference.\(^{22}\)

In the empirical applications, I use three implications. First, side-compatibility is high for a principal that engages multiple vendors, because this creates alternative vendors each with access to economies of scope (given their sunk partnership-specific fixed costs). Second, side-compatibility is low when a seller dedicates most of its capacity to a single buyer because it cannot then sell adapted trades alongside the basic trade. Third, side-compatibility is trivial if the contract has exclusivity clauses prohibiting the relevant side trades.

**Investment categories.** In the introductory illustration, IBM’s specific investment (cost-cutting) has $\gamma \simeq 0$, $\kappa = 0$ and $\sigma > 0$ so $\psi(= \sigma) > \gamma$, while Amex’s adaptation investment has $\kappa < \gamma$ and $\sigma = 0$ so $\psi = \kappa < \gamma$. Long-term contracting raises IBM’s default returns on cost-cutting (for which $\psi > \gamma$), but lowers Amex’s default returns on investing in adaptations (for which $\psi < \gamma$). To generalize this distinction between the contract’s protection and market-

\(^{19}\) There are two possible reasons why the adapted contract ($Z$) cannot be written at stage 0. Either the adapted trade had not been discovered or because the traders could not write the necessary contractual modification in advance.

\(^{20}\) In general, the adaptation effect equals one minus the sum of the self and cross effects (see section 4).

\(^{21}\) The novelty of my model is in allowing, and analyzing when $\psi < \gamma$ (this is why contracting reduces incentives). I say that side-compatibility is complete if $\kappa = \gamma$. $\kappa > \gamma$ only occurs in the special case where competitors can provide an adaptation that complements a basic trade which they cannot provide. Note that $\kappa < \gamma$ implies

\(^{22}\) Side-compatibility is high when adaptations require additional trades that are relatively unrelated to the basic trade, because the relevant economies of scope are then lower.
shielding effects, I define an investment as **type 1** if it has \( \psi > \gamma \) and as **type 2** if it has \( \gamma > \psi \) (see also section 4 extensions). Each investor, \( j \), makes one investment of each type - it is straightforward to generalize to any number of investments. I denote \( j \)'s type 1 and type 2 investments by \( e_j \) and \( i_j \), respectively. The investments have additively separable returns, \( W_j(e_j) \) and \( V_j(i_j) \), satisfying standard concavity, monotonicity and Inada boundary conditions that guarantee interior investment equilibria.

**Assumption 1.** For \( j \in \{ P, A \} \),

\[
W'_j(e_j) > 0, W''_j(e_j) < 0, \text{ on } e_j \geq 0, \text{ and } \lim_{e_j \to 0^+} W'_j(e_j) = \infty, \lim_{e_j \to \infty} W'_j(e_j) = 0;
\]

\[
V'_j(i_j) > 0, V''_j(i_j) < 0, \text{ on } i_j \geq 0, \text{ and } \lim_{i_j \to 0^+} V'_j(i_j) = \infty, \lim_{i_j \to \infty} V'_j(i_j) = 0.
\]

In sections 2 and 3 only, I also restrict the parameters \( \gamma \) and \( \psi \) to the unit interval, \( [0,1] \).

**Assumption 2.** For each investment, \( \gamma, \psi \) lie in \( [0,1] \).

In my base case, all payoffs are additively separable in costs, benefits and transfers across time. \( P \) and \( A \) are risk-neutral\(^{23} \) and face no wealth constraints. I let \(-j\) denote \( P \) when \( j = A \) and \( A \) when \( j = P \), and I suppress the statement, “for \( j = P, A \).” Investment costs and trade payoffs are in money metric units and I normalize to the case with no time-discounting. So \( P \) and \( A \)'s overall objectives are given by the utility functions, \( U_j \equiv u_j - e_j - i_j + T_j \) where \( T_j \) is the net transfer from \(-j\) to \( j \) and \( u_j \) denotes the stage 3 trade payoffs,

\[
u_j \equiv \begin{cases} 
\gamma e_j \cdot W_j(e_j) + \gamma i_j \cdot V_j(i_j), & \text{if } x, x' \\
\psi e_j \cdot W_j(e_j) + \psi i_j \cdot V_j(i_j), & \text{if } x, x' \\
W_j(e_j) + V_j(i_j), & \text{under optimal trading}
\end{cases}
\]

I sketch the timing for this section in Figure 1.

<table>
<thead>
<tr>
<th>Stage 0</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negotiate</td>
<td>Invest</td>
<td>Renegotiate</td>
<td>Trade</td>
</tr>
<tr>
<td>( X ) or ( \Phi )</td>
<td>( e_P, i_P, ) and ( e_A, i_A )</td>
<td>( Z, X ) or ( \Phi )</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1: Timeline for Base Model**

\(^{23}\)I follow Goldberg and Erickson (1987) in ignoring risk. Recent evidence suggests risk-sharing often plays a limited role in contract design even in the classic case of small farmers (see Allen and Lueck, 1995, 1999).
The first-best. In the first-best, P and A can fix e and i cooperatively at stage 0. In stage 3, the trade surplus is \( u_P + u_A = \sum_{j=P,A} (W_j(e_j) + V_j(i_j)) \). So in stage 0, they choose e and i to maximize their total surplus,\(^{24}\)

\[
U_P(e, i) + U_A(e, i) = \sum_{j=P,A} [W_j(e_j) + V_j(i_j) - (e_j + i_j)]
\]

Assumption 1 ensures that the first order conditions are both necessary and sufficient, and give unique solutions to the first-best which I denote with a superscript*,

\[
V_j'(i_j^*) - 1 = 0 \quad \text{and} \quad W_j'(e_j^*) - 1 = 0 (1)
\]

In the second-best equilibrium, P and A choose e and i non-cooperatively, because these investments, their costs, and their resulting payoffs, u, are all non-verifiable (e.g., outsiders cannot measure the quality or cost of IBM’s coaching investments). I therefore compare the Subgame Perfect Equilibria from alternative feasible stage 0 contracts to derive the optimal contract. I begin (in 2.1 and 2.2) with the two extreme contracts where P and A choose either X or \( \Phi \) at stage 0. Superscripts LTC and STC indicate the equilibrium values from long-term contracting (choosing X) and short-term contracting (choosing \( \Phi \)), respectively. I then treat the general case in 2.3.

2.1 Equilibrium with long-term contracting.

When P and A agree on X at stage 0, their default payoffs in stage 2 renegotiation are given by the expression, \( \psi_{e_j} \cdot W_j(e_j) + \psi_{i_j} \cdot V_j(i_j) \). The total gain available on renegotiation is therefore,

\[
\sum_{j=P,A} \left[ (1 - \psi_{e_j}) \cdot W_j(e_j) + (1 - \psi_{i_j}) \cdot V_j(i_j) \right]
\]

P and A each get their default payoffs plus half of these renegotiation gains, so after making investments e and i, their expected payoffs from stage 2 onwards are given by,

\[
\frac{1}{2} (1 + \psi_{e_j}) W_j(e_j) + \frac{1}{2} (1 + \psi_{i_j}) V_j(i_j) + K(e_{-j}, i_{-j})
\]

where \( K(e_{-j}, i_{-j}) \equiv \frac{1}{2} \left[ (1 - \psi_{e_{-j}}) \cdot W_{-j} + (1 - \psi_{i_{-j}}) \cdot V_{-j} \right] \). I solve for the Subgame Perfect Equilibrium using backward induction. Party j invests to maximize this expected return less its investment cost, \( e_j + i_j \). Since \( K(\cdot) \) does not depend on \( e_j \) and \( i_j \), j’s problem is,

\[
\max_{e_j, i_j} \left\{ \frac{1}{2} (1 + \psi_{e_j}) W_j(e_j) + \frac{1}{2} (1 + \psi_{i_j}) V_j(i_j) - (e_j + i_j) \right\}
\]

\(^{24}\)T_P and T_A (summing to 0) depend on relative bargaining power and participation constraints.
The first order conditions are again necessary and sufficient,

\[
\frac{1}{2} \left( 1 + \psi_{ij} \right) W'_j (e_j) - 1 = 0
\]

\[
\frac{1}{2} \left( 1 + \psi_{ij} \right) V'_j (i_j) - 1 = 0
\]

When \( \psi = 1 \), these conditions replicate the equation (1) of the first-best, but for \( \psi < 1 \), concavity of \( W \) and \( V \) implies underinvestment. Investor \( j \)'s investments increase with \( j \)'s share of default returns,\(^{25}\) which increase with the contract-compatibility parameters, \( \psi_{e_j} \) and \( \psi_{i_j} \). The general intuition is familiar: A higher investment return in the default outcome reduces dependence on negotiating with the key trading partner. The investor therefore loses a smaller share of investment returns in renegotiation, and better internalizes the investment.

**Proposition 1.** \( e_{j}^{LTC} \) rises with \( \psi_{e_j} \) and \( i_{j}^{LTC} \) rises with \( \psi_{i_j} \). Assumption 2 implies underinvestment in \( e_P \) and \( e_A \) and weak underinvestment in \( i_P \) and \( i_A \): \( i_{j}^{LTC} < i_{j}^{*} \) and \( e_{j}^{LTC} \leq e_{j}^{*} \) with equality only when \( \psi_{e_j} = 1 \).

In words, long-term contracting protects investments to the extent that they are contract-compatible. Since \( \psi = \sigma + \kappa \), this captures how the contract protects an investment if directly compatible (\( \psi \) is increasing in \( \sigma \)), and permits motivation by market forces if the investment is side-compatible (\( \psi \) is increasing in \( \kappa \)).

### 2.2 Equilibrium with short-term contracting

When \( P \) and \( A \) choose \( \Phi \) at stage 0, their stage 2 default payoffs are given by \( \gamma_{e_j} \cdot W_j (e_j) + \gamma_{i_j} \cdot V_j (i_j) \). So, by analogy with the long-term contracting case, the first order conditions become,

\[
\frac{1}{2} \left( 1 + \gamma_{e_j} \right) W'_j (e_j) - 1 = 0
\]

\[
\frac{1}{2} \left( 1 + \gamma_{i_j} \right) V'_j (i_j) - 1 = 0
\]

So investments now increase with \( \gamma \) instead of \( \psi \), and there is underinvestment when \( \gamma < 1 \).

**Proposition 2.** \( e_{j}^{STC} \) rises with \( \gamma_{e_j} \) and \( i_{j}^{STC} \) rises with \( \gamma_{i_j} \). Assumption 2 implies underinvestment in \( e_P \) and \( e_A \) and weak underinvestment in \( i_P \) and \( i_A \): \( e_{j}^{STC} < e_{j}^{*} \) and \( i_{j}^{STC} \leq i_{j}^{*} \) with equality only when \( \gamma_{i_j} = 1 \).

\(^{25}\)As always, a higher default return implies lower dependence on negotiating with the original trade partner. The investor loses a smaller share of investment returns in renegotiation, and better internalizes its investment.
In words, short-term contracting allows market forces to motivate investments to the extent that they are general. This is simply a positive statement of the classic holdup problem - specificity causes underinvestment. The intuition again follows from asking whether an investor can appropriate investment returns without having to renegotiate. Under short-term contracting, the default returns are determined by the investor’s market alternatives reflected in $\gamma$. The problem is that market forces do not motivate specific investments.

2.3 Equilibrium with intermediate contract lengths

In Section 5, contracts can extend over multiple trading periods, but even in the basic model, I can treat contract length as a continuous variable by allowing for stochastic enforcement. I let $\alpha$ denote the probability that $X$ is enforced (in default of renegotiation). With converse probability $1 - \alpha$, $P$ and $A$’s default contract is $\Phi$. I assume that $P$ and $A$ can choose any $\alpha \in [0, 1]$ at stage 0 - for instance, by varying contractual ambiguity or breach damages (see 6.2). This generalizes the above analysis because $\alpha = 0$ corresponds to short-term contracting ($STC$) and $\alpha = 1$ corresponds to long-term contracting ($LTC$). I refer to $\alpha$ as contract length - the contract enforces trade for on average $\alpha \cdot G$ time units where $G$, the time elapsing between stages 0 and 3, is the common gestation period of all investment. The interpretation of $\alpha$ as a deterministic contract length is validated in the multi-period model of section 5.

The default payoffs are now convex combinations of those from $X$ and $\Phi$ with weights, $\alpha$ and $1 - \alpha$, respectively. The first-order conditions are therefore,

$$\frac{1}{2} \left( 1 + \alpha \cdot \psi_j + (1 - \alpha) \gamma_j \right) W_j'(e_j) = 1$$
$$\frac{1}{2} \left( 1 + \alpha \cdot \psi_i + (1 - \alpha) \gamma_i \right) V_j'(i_j) = 1$$

This set of equations allows me to generalize the first two propositions and show how contract length affects investment incentives as a function of the sign and size of $\psi - \gamma$. Raising $\alpha$ shifts weight from $\gamma$ onto $\psi$. This raises $F \equiv 1 + \alpha \cdot \psi + (1 - \alpha) \gamma$ when $\psi > \gamma$, and lowers it when $\psi < \gamma$. Equations (2) then imply that increasing $\alpha$ increases $W_j'(e_j)$ and decreases $V_j'(i_j)$, so $e_j$ rises and $i_j$ falls (as $W$ and $V$ are concave).

**Proposition 3.** (a) Increasing contract length raises type 1 and lowers type 2 investments. Mathematically, $\frac{de_j(\alpha)}{d\alpha} > 0$, $\frac{di_j(\alpha)}{d\alpha} < 0$. (b) Each investment rises with $\gamma$ (strictly if $\alpha < 1$) and with $\psi$ (strictly if $\alpha > 0$). (c) With Assumption 2, underinvestment is the only possible inefficiency.

$$e_j^{STC} < e_j(\alpha) < e_j^{LTC} \leq e_j^*$$ and $$i_j^{LTC} < i_j(\alpha) < i_j^{STC} \leq i_j^*$$ for all $\alpha \in (0, 1)$

I can redefine $\alpha$ to be $P$ and $A$’s prior estimate of whether stage 3 trade will be induced.
This captures the key tradeoff between increasing contract length to protect type 1 investments and reducing contract length so that market forces can better motivate type 2 investments. Type 1 investments, \(e_P\) and \(e_A\), are specific but contract-compatible (\(\gamma < \psi\)) so the contract protects them better than do market forces. Type 2 investments, \(i_P\) and \(i_A\), are general and contract-incompatible (\(\gamma > \psi\)) so for them, market forces are better than the long-term contract (with its side-compatible market forces). In the next section, I predict contract length by trading off contractual protection of \(e\) against its cost in shielding out contract-incompatible market forces that reward \(i\).

3 Optimal Contract Length in the Basic Model

\(P\) and \(A\) choose contract length at stage 0. In this section, I show how optimal length is determined by the relative importance of different investments and the effectiveness and compatibility of contracts and market forces. In stage 0 negotiations, \(P\) and \(A\) can use up-front transfers to share any gains from a surplus increasing contract. So they choose \(\alpha\) to maximize their total surplus, subject (unlike in the first-best) to the “incentive compatibility” conditions (2) on \(e\) and \(i\). I denote the unique solutions of (2) by \(e_j = e_j(\alpha)\) and \(i_j = i_j(\alpha)\). So \(P\) and \(A\)’s problem is,

\[
\max_{\alpha} \sum_{j=P,A} \left( W_j(e_j(\alpha)) + V_j(i_j(\alpha)) - (e_j(\alpha) + i_j(\alpha)) \right)
\]

(3)

Propositions 1-3 show that underinvestment is the only efficiency problem in the basic model. Proposition 3 also shows how to motivate more investment: increase \(\alpha\) to improve type 1 investments, \(e\), and decrease \(\alpha\) to improve type 2 investments, \(i\). Intuition therefore suggests that \(\alpha\) should increase with the importance of \(e_P\) and \(e_A\) relative to \(i_P\) and \(i_A\). To formalize this idea, I scale up the productivity of investments \(e_j, i_j\) by the importance parameters, \(E_j, I_j > 0\) - i.e., I replace \(W_j(e_j)\) by \(E_j \cdot W_j(e_j)\) and \(V_j(i_j)\) by \(I_j \cdot V_j(i_j)\). \(P\) and \(A\)’s problem is then,

\[
\max_{\alpha} \sum_{j=P,A} \left( E_j \cdot W_j(e_j(\alpha, E_j)) + I_j \cdot V_j(i_j(\alpha, I_j)) - (e_j(\alpha, E_j) + i_j(\alpha, I_j)) \right)
\]

(4)

where \(e_j(\alpha, E_j)\) and \(i_j(\alpha, I_j)\) solve the adjusted incentive compatibility conditions,

\[
\frac{1}{2} F(\alpha; \psi e_j, \gamma e_j) E_j \cdot W'_j(e_j) = 1
\]

\[
\frac{1}{2} F(\alpha; \psi i_j, \gamma i_j) I_j \cdot V'_j(i_j) = 1
\]

where \(F(\alpha; \psi, \gamma) \equiv (1 + \alpha \cdot \psi + (1 - \alpha) \gamma)\)

(5)

I make two assumptions to deal with two minor complications.
Assumption 3. \[ W_j^*(e_j) W_j''(e_j) > \left( W_j''(e_j) \right)^2 \text{ and } V_j^*(i_j) V_j''(i_j) > \left( V_j''(i_j) \right)^2, \forall e_j, i_j. \]

Assumption 4. Maximand (4) is concave. [See Appendix for sufficient conditions.]

Assumption 3 (a regularity condition familiar from insurance contracting\textsuperscript{27}) guarantees domination of an effect (namely, that raising an investment’s importance raises the incentive to invest for fixed \( \alpha \) - e.g., \( \frac{\partial e_j}{\partial E_j} > 0 \)) that countervails against the need to use \( \alpha \) to raise incentives. Assumption 4 ensures a unique solution and I can prove the intuitive result as follows,

**Proposition 4.** Under assumptions 3 and 4, optimal contract length increases with the importance of investments of type 1, and decreases with importance of type 2’s: \( \alpha (E, I) \) has \( \frac{\partial \alpha}{\partial E_j} > 0 \) and \( \frac{\partial \alpha}{\partial I_j} < 0 \).

This result is consistent with empirical evidence (see e.g., Brickley et al., 2003) on the positive correlation between contract length and the importance of non-contractible specific investments, \( E_A \). Empiricists use the size of contractible specific investments as a proxy for the importance of non-contractible specific investments (\( E_A \)). This proxy is imperfect but reasonable, because the two types of investment are strongly complementary. For instance, when IT outsourcing deals involve a significant (contractible) transfer of workers and assets from client to vendor, the complementary investments in retraining and reorganization are largely specific and non-contractible (as explained above for IBM’s coaching). Using transfer size as a proxy, proposition 4 readily explains why almost all the longer IT contracts (those over 5 years long) occur in these “merger and acquisition” type deals.

The effectiveness of markets and contracts (captured by \( \gamma \) and \( \psi \)) also affects optimal contract length. There are two types of effect for any given investment. First, if \( \psi \) and \( \gamma \) change so that \( \delta = \psi - \gamma \) rises while \( F (\alpha) \equiv 1 + \alpha \cdot \psi + (1 - \alpha) \gamma \) is fixed, long-term contracting becomes more effective relative to the market forces freed by short-term contracting. This intuitively favors the use of longer contracts and should increase \( \alpha \). I call this a substitution effect, because \( P \) and \( A \) substitute market forces for contract length according to their relative effectiveness as investment motivators. Second, an increase in \( \gamma \) or \( \psi \) directly raises the investment level for any given \( \alpha \). This level effect is determined by changes in \( F (\alpha) \) for fixed \( \alpha \) and \( \delta \). The level effect reduces the need to adjust \( \alpha \) in favour of the investment, so the level effect on \( \alpha \) is negative for type 1 investments and positive for type 2 investments.

\textsuperscript{27}There is no precautionary saving interpretation here, but consider the sharp countervailing effect of a fixed cost self-investment: when the investment’s importance generates incentives exceeding the fixed cost, \( \alpha \) can be decreased. Assumptions 1 and 3 rule out generalized versions of this problem.
Proposition 5. Contract length rises when type 1 investments become more specific ($\gamma$ falls) and falls when type 2 investments become less contract-compatible ($\psi$ falls).

In general: Changes that increase $\delta = \psi - \gamma$ have a positive “substitution effect”: $\frac{\partial \alpha(\gamma, \psi)}{\partial \delta}| F = \hat{F} > 0$; Changes that increase $F(\alpha; \psi, \gamma)$ have a “level” effect that is negative for type 1, but positive for type 2, investments: $\frac{\partial \alpha(\gamma, \psi)}{\partial F}|_{\delta = \hat{\delta}} \text{sign} = -\hat{\delta}$.

I use two direct corollaries in section 6. First, contract length rises with the side-compatibility of type 2 investments, because a rise in $\kappa$ implies a rise in $\psi = \sigma + \kappa$. Given that side-compatibility tends to be higher in multi-sourcing, this generates a tendency for longer contracts in multi-vendor situations. Second, contract length rises when market alternatives fall, because this makes investments more specific, i.e., $\gamma$ falls. This claim, famously supported by Joskow (1987) and others, is clear for type 1 investments. My model shows that there is a complication for type 2 investments. When type 2 investments become more specific, the level effect could motivate shorter contracts. This is most likely dominated by the substitution effects (which are positive for all investments) and the positive level effects (from lowering $\gamma$) on type 1 investments. Furthermore, investments could switch from type 2 to type 1 so that the relative importance of type 1’s increases, and contracts are longer by proposition 5.

4 Cross Effects, Waste and Bargaining Asymmetry

This section extends the model to allow for investment externalities, wasteful investment and bargaining asymmetries. The cross effect is the investment externality that occurs under the basic contract $X$. A cross-investment is defined by having a positive cross effect. This occurs whenever the basic contract induces a (non-contractible) quality of trade that depends on prior investment by the trading partner. In the introductory example with the pay-as-you-use contract, IBM’s storage efficiency affects trade quality, because Amex then gets more effective server space per dollar. Cross effects may be negative. For instance, IBM might be able to cut its cost at some expense to quality (as occurs under privatization in Hart et al., 1997, when contracts cannot define quality). In this case, IBM’s private return may exceed the social return on its investment. So I allow $\psi > 1$. $\psi > 1$ also captures the fact that specific investments are of no use if it is optimal to separate (i.e., when $Z$ involves separate trading with some probability). Similarly, $\gamma > 1$ is possible when investments (e.g., in advertising or search)

---

28 E.g., for adaptation investments. Note that raising $\psi$ on any investment raises contract length if level effects are limited - as guaranteed in the case $\alpha = 0$.

29 A fixed price contract prevents this cross effect, but then Amex’s varying server needs have a cross-effect on IBM, and IBM can holdup Amex’s adaptations (see 6.1).
generate alternatives that are mostly used as threat points. I also allow for investments to be wasteful - in that they have no (or negative) social return. For instance, investing to make a credible threat (e.g., exploiting a contractual loophole) that is never carried out because of efficient renegotiation. I define the cross-general effect as the externality occurring under \( \Phi \). This is the market-based “cross effect” occurring when one party’s investment improves the other party’s market alternatives under short-term contracting. For instance, Amex’s market alternatives improve if Amex learns technical or marketing knowledge from watching IBM invest in their joint trade. Conversely, suppliers may learn from their buyers - indeed many contract-manufacturers have become own-brand producers.\(^{30}\) To complete the picture, I allow for negative cross-general effects, as when an investment harms the trade partner’s market reputation.

It is simplest to extend the model, derive the general results and then explain the intuitive impacts on contract length. I extend the model by allowing \( j \)'s investment, \( e_j \), to raise \(-j\)'s payoff by \( \psi_{e_j}^{\text{cross}} \cdot W_j (e_j) \) under \( X \) (the cross effect), and by \( \gamma_{e_j}^{\text{cross}} \cdot W_j (e_j) \) under \( \Phi \) (the cross-general effect). I also allow \( e_j \) to be a wasteful investment by removing the additive \( W_j (e_j) \) term from the trade surplus under \( Z \).\(^{31}\) Similarly, for \( i,j \). The impact of \( j \)'s positive cross effects is to reduce \( j \)'s payoff from renegotiation, because \(-j\)'s default payoff is increased. The first order conditions (5) are unchanged except that (a) \( \psi \) and \( \gamma \) are replaced by \( \tilde{\psi} \equiv \psi - \psi^{\text{cross}} \) and \( \tilde{\gamma} \equiv \gamma - \gamma^{\text{cross}} \), and (b) in the case where the investment is wasteful, \( F (\alpha; \psi, \gamma) = 1 + \alpha \cdot \psi + (1 - \alpha) \gamma \) is replaced by \( G (\alpha; \tilde{\psi}, \tilde{\gamma}) \equiv \alpha \cdot \tilde{\psi} + (1 - \alpha) \tilde{\gamma} \). The natural extension of the type 1 investment is defined by \( \tilde{\delta} (\equiv \tilde{\psi} - \tilde{\gamma}) > 0 \) while type 2 investments are defined by \( \delta < 0 \). The implications of the cross effects in the basic model are immediate corollaries of propositions 1 to 5 because all these results continue to hold for \( \tilde{\psi} \) and \( \tilde{\gamma} \) (when less than unity). The cross effects simply counteract against the corresponding self effects. Rather than rewrite the propositions, I describe the extreme cases and some minor subtleties.

An investment increases with contract length if and only if \( \psi - \psi^{\text{cross}} > \gamma - \gamma^{\text{cross}} \), i.e., \( \delta > 0 \). A pure cross-investment is one for which \( \psi^{\text{cross}} > 0 \), while the other parameters are zero. This is a type 2 investment because \( \delta = -\psi^{\text{cross}} < 0 \). So it falls with contract length, \( \alpha \). The problem is not one of market-shielding. After all, \( \gamma = 0 \). Here instead, the long-term contract is costly (as in Che and Hausch, 1999), because it increases the investment’s positive externality. Proposition 4 states that contracts should become shorter as cross-investments become more important, but there is one complication. Intermediate breach penalties might now be

\(^{30}\)See Arruñada and Vázquez (2004) and Lee and Hoyt (2001), on how Solectron began to sell its own-brand products after working for IBM, HP and Mitsubishi.

\(^{31}\)If \( e_j \) actually reduces this social return, the effects are simply more pronounced.
optimal. Concretely, if only one trader makes cross- and self- investments, the breach penalty of the non-investing trader could be reduced so that it credibly threatens to terminate after low cross-investment. Such ‘option contract’ solutions also reduce $\alpha$ below 1 in generic (stochastic) settings, and are not always robust, so I defer a general analysis.\footnote{Che and Hausch (1999) argue that such option schemes do not work (given renegotiation), because the trader would threaten to terminate even after high cross-investment. Ellman (2003) shows that this critique is invalid when options are decided by a trading decision (see also Watson, 2003), while recognizing that reputation effects can reintroduce the problem.}

A pure cross-general investment is one where only $\gamma_{cross} > 0$. This is a type 1 investment since $\delta = \gamma_{cross} > 0$, so it rises with $\alpha$. The contract helps because it shields out the market externality.\footnote{Note that $\psi_{cross}$ would be positive if $\kappa_{cross} > 0$ since $\psi_{cross} = \sigma_{cross} + \kappa_{cross}$ (natural notation extension).}

Proposition 4 predicts longer contracts when cross-general effects are important. A contract imposing exclusivity alone may (if legal) be more effective in preventing cross-general externalities (see Segal and Whinston, 2000), but performance contracting often dominates by motivating self-investments and side-compatible adaptations.

Asymmetries in bargaining do not change the nature of these effects, but they do change the relative importance of cross and self effects, because self effects lead to a benefit without need for bargaining power, while cross effects only matter through the traders’ renegotiation shares. If $j$ now wins a share $\theta_j \in [0, 1]$ of the renegotiation returns, $j$’s incentive conditions are as in (5) except that (a) $\bar{\psi}$ and $\bar{\gamma}$ replaced by $\psi(\theta) \equiv (1 - \theta) \psi - \theta \psi_{cross}$ and $\gamma(\theta) \equiv (1 - \theta) \gamma - \theta \gamma_{cross}$ and (b) $F(\alpha; \psi, \gamma)$ is replaced by $H(\alpha; \psi, \gamma, \theta) \equiv 2[\theta + \alpha \cdot \psi(\theta) + (1 - \alpha) \gamma(\theta)]$. The direction of effects is as before, once the definition of investment type is extended: an investment is of type 1 if $\delta(\theta) > 0$ where $\delta(\theta) \equiv \psi(\theta) - \gamma(\theta)$, and type 2 if $\delta(\theta) < 0$.

When an investment is wasteful, the impact of contract length is determined by its parameters $\bar{\psi}$ and $\bar{\gamma}$ exactly as for productive investments apart from the level effect implicit in $G = F - 1$ (for the case with generalized bargaining powers, $G \equiv H - \theta$). However, the maximand (4) is only extended by subtraction of its cost, since there is no social return. So $P$ and $A$ aim to minimize the cost, and the message of proposition 4 is exactly inverted as shown in proposition 6 for a wasteful investment, $k_j$, with private returns proportional to $\Omega \cdot B_j(\kappa_j)$ ($B_j$ must satisfy the above regularity conditions, except assumption 3 is unnecessary as $\Omega$’s level effect is lower).

**Proposition 6.** Optimal contract length decreases with the importance of wasteful type 1 investments, and increases with the importance of wasteful type 2 investments: $\frac{d\alpha}{d\Omega} < 0$ if and only if $\delta > 0$.  

When $\bar{\psi} > 1$ (either from negative cross-investment effects or from $\psi > 1$ or both), there
is a risk of over-investment - for any $\alpha > \frac{1 - \bar{\psi}}{\psi - \bar{\gamma}}$. So, for high $\alpha$, changing the productivity (importance) of this investment has the same effect as if the investment were wasteful, while for low $\alpha$, the implications are as in proposition 4. When $\bar{\gamma} > 1$, this problem is inverted.\(^{34}\)

In conclusion, the impact of cross effects ($\psi^{\text{cross}}$ and $\gamma^{\text{cross}}$) is the inverse of the corresponding self effects ($\psi$ and $\gamma$) and the impact of importance on contract length is inverted for an investment that is wasteful or excessive. When traders are inexperienced or uncertainty and complexity are high, it is harder to write contracts that pin down quality, so there is a greater risk of cross effects. This section helps explain why contracts are often short in such settings: contract length is reduced to better motivate cross-investments and to reduce over-investment in investments with negative cross effects. Uncertainty also increases the importance of adaptations which also explains the shorter contracts.

5 Temporal Extension

This section analyzes multi-period extensions of the trading model. Stage 3 trading is spread over time\(^ {35}\) and is now subdivided into $N$ discrete trade decisions, each lasting $l$ units of time. A simple long-term contract enforcing trade in the first $m \leq N$ substages has length $L = m \cdot l + G$. Increasing $L$ protects self-investments for longer, but also shields out for longer the market forces that protect adaptations. Is this tradeoff identical to that for $\alpha$ in the basic model? When do optimal contracts take this simple form?

Retaining the standard assumption that parties can always renegotiate, I need to add $N$ renegotiation stages - one before each trading decision. In the simplest extension, there is still only one investment stage and no history dependence within the extended trading interval, so the trade payoffs in each substage are scalar multiples of the trade payoffs from the basic model. The impact of the stage 1 investments may vary with time, so I let the compatibility, generality and importance parameters depend on $n$. The overall game is as before except that stages 2 and 3 are replaced by their $N$-fold replication and the initial contract determines a probability $\alpha_n$ of trade enforcement via $X_n$ (equivalent to $X$) for each of the $n$ trade stages. The extended timing is therefore: (Stage 0) P and A negotiate $(\alpha_n)_{n=1}^N$ and lump-sum transfers; (Stage 1) P and A invest; (Stage 2) Renegotiation over $X_n$; (Stage 2 + 1) $n$'th trading decisions $(\alpha_n)_{n=1}^N$.

\(^{34}\)To complete the generalization, $\psi$ and $\gamma$ might also be negative, but this has no special effect other than possible corner solutions at zero investment.

\(^{35}\)Trade is often spread over time because production is time-intensive, being limited by capacity and procedural constraints. For services like IT, it is because the buyer's needs are spread over time. See 6.1 on endogenizing trade intensity.
Given the absence of history-dependence within the subgame starting from stage 2, the continuation payoffs equal the sum of the equilibrium payoffs from the \( N \) paired stages \((2n \text{ and } 2n+1)\)\(\)\(n=1\ldots N\). The first-order conditions therefore modify (5) into,

\[
\left( \sum_{n=1}^{N} \frac{1}{2} F (\alpha_n; \tilde{\psi}_{e,j,n}, \tilde{\gamma}_{e,j,n}) \cdot E_{j,n} \right) \cdot W_j' (e_j) = 1,
\]

\[
\left( \sum_{n=1}^{N} \frac{1}{2} F (\alpha_n; \tilde{\psi}_{i,j,n}, \tilde{\gamma}_{i,j,n}) \cdot I_{j,n} \right) \cdot V_j' (i_j) = 1
\] (6)

In the case of time invariant technology, the summations in (6) simplify to

\[
\frac{N}{2} F (\alpha_n; \tilde{\psi}_{i,j,n}, \tilde{\gamma}_{i,j,n}) \cdot E_j \quad \text{and} \quad \frac{N}{2} F (\alpha_n; \tilde{\psi}_{i,j,n}, \tilde{\gamma}_{i,j,n}) \cdot I_j,
\]

where \( \alpha = \sum_{n=1}^{N} \alpha_n \). So contracts with the same \( \alpha \) are efficiency equivalent. In particular, the simple long-term contract of length \( L = m \cdot l \) has \( \alpha_n = 1_{(n \leq m)} \), so its \( \alpha = \frac{m}{N} = \frac{L}{L} \) (where \( L = N \cdot l \) is the maximal trade duration, and I assume \( G \) is trivial) and I can state,

**Proposition 7.** In the multi-period extension with time invariance, restricting to simple long-term contracts has no efficiency cost and propositions 1-6 all hold with \( L \) replacing \( \alpha \).

Simple contracts are optimal if self-investments become redundant over time and/or adaptation investments become more important over time, because simple contracts crowd trade enforcement into the earliest substages where contract protection is most effective relative to market forces.\(^{37}\) History dependence within the trade interval generates a pair of arguably stronger reasons for simple contracts. First, switching to an alternative partner for the main trade may reduce the marginal value of specific investments when returning to joint trade. For instance, switching may require reorganizations that undo the specific investments. The benefit from imposing performance after a gap (in enforcement of joint trade) is then reduced, so it is better to avoid the performance gap by using a simple contract. Second, fixed costs of switching reduce the attractiveness of market alternatives for a trader with plans to switch

---

\(^{36}\)Renegotiation and interdependence among the \( \alpha_n \) could create history-dependence inside the subgame, but the game with renegotiation restricted to adjusting the contract only for the upcoming trading period and independent \( \alpha_n \) has the same equilibria (see Ellman, 2003).

\(^{37}\)The general condition is that the importance of type 1 investments grow at a lower rate than for type 2’s - i.e., \( k_n \equiv \frac{E_{j,n+1}}{E_{j,n}} < K_n \equiv \frac{I_{j,n+1}}{I_{j,n}}, \forall n, \forall j \). The proof is simple: Suppose that \( \alpha_{n+1} > 0 \) when \( \alpha_n < 1 \). Decreasing \( \alpha_{n+1} \) by (any feasible) \( \varepsilon > 0 \) and increasing \( \alpha_n \) by \( K_n \cdot \varepsilon \) fixes the incentive on all type 2 investments and increases the incentives for all type 1 investments by \( (K_n - k_n) \frac{\varepsilon}{2} \delta > 0 \). This improvement proves that \( \hat{\alpha}_n < 1 \Rightarrow \hat{\alpha}_{n+1} = 0 \) \( \forall n \in \{1, 2, \ldots, N\} \) so gaps are always avoided.
back to joint trade after a performance gap. When switching costs are high, market alternatives cease to be credible during a gap, so market-shielding is not escaped and it is better to fill any gap with performance. Again, simple contracts are uniquely optimal and proposition 7 can be strengthened, because the restriction to simple contracts is superfluous.

Setting an investment’s productivity parameters to zero for early substages allows me to represent the gestation period before it becomes productive (by the time elapsed between stage 1 and the first substage in which the investment yields a return). Two implications are immediate. First, \( L \) should never be less than \( \min_j G_{ij} \), because the contract has no market-shielding cost until type 2 investments become productive. Second, if \( L > 0 \), then \( L \) should exceed \( \min_j G_{eij} \), because otherwise the contract has no protection benefit. This may help explain why agriculture contracts are longer in the case of fruit trees as shown in Bandiera’s (2002) historical data. However, when \( G_{eij} \) gets too large, the market-shielding cost may be prohibitive and \( L \) will fall back, as traders abandon the idea of protecting \( e_j \). Further implications depend on the time profiles of investment productivity and can be analyzed using my extended model.

Investments are often made during trade. P and A may trade during stage 1 in my model, but I have not allowed for ongoing investments. Introducing additional investment stages after stage 2 requires ongoing renegotiation - the new contract at each stage must be designed to optimize incentives for the upcoming investment problem. My framework would then predict the length of each new contract, assuming intertemporal additivity for all investments. Guriev and Kvassov (2004) study precisely this problem in the case of a single investment, though without the market-shielding problem, and under the premise that renegotiation is exogenously costly. Assuming that the length of the IBM-Amex contract was mostly driven by the need

---

38 In fact it may be more severe than during performance contracting, because adaptation trades with low switching costs are usually useless without a basic trade.

39 There is a complication when allowing for intermediate breach penalties. Large switching costs turn alternative trades into “outside options,” as defined in Osborne and Rubinstein (1990) and this generates multiple equilibria. Applying the Outside Option Principle as in MacLeod and Malcomson (1993), generates identical results if outside options never bind, but when one party’s outside option binds, the other party gets the full share of renegotiation surplus. In the extreme case where only one side makes relationship-specific investments, the tradeoff between protection and market forces could be escaped by giving an attractive breach option to the party without any specific investment.

40 Choosing \( \Phi \) then means choosing a short-term contract enforcing joint trade in stage 1 - and choosing \( X \) implies imposing joint trade in stage 1 as well as in the post-investment trading periods.

41 They show that fixed term contracts must be continuously renegotiated to motivate ongoing investment, whereas contracts specifying a minimum advance notice for termination (called “ever-greens”) provide ongoing investment incentives without need for continuous renegotiation. Adding my market-shielding effect to their problem would lead to shorter advance notice requirements. Che and Sákovicz (2004) also allow for ongoing
to motivate early investments in reorganization and retraining, one might have expected renegotiation to a shorter contract once those costs were sunk. I suspect that this did not happen for two reasons. First, the remaining length of the contract and size of breach penalties anyway fall automatically with time. Second, asymmetric information may frustrate renegotiations: If Amex does not observe IBM’s investments perfectly, it would rightly fear that IBM would only agree to shorten the contract when its self-investments are relatively low.

I conclude this section with a brief comment on measuring contract length. The simple contracts motivated above are intuitive, but measurement remains non-trivial when breach is stochastic (as in 6.1). The “effective” length of a contract (defined in this paper as the amount of time over which a single performance contract would induce trade in the absence of renegotiation) can, through premature breach, be lower than the “nominal” length reported in the contract. Nonetheless, reported nominal length (7 years in the case of IBM-Amex) is a good proxy for effective length, because breach penalties often fall sharply to zero at the end of the reported period.43

6 Contracts, Endogenous Side-Compatibility and the Empirical Puzzles

This section describes how common features of actual contracts and the organization of trade affect side-compatibility. I consider quantity variation, menu contracts and ‘flexible’ contracts, breach penalties and informal enforcement, multi-sourcing and selective outsourcing, capacity dedication and contractual restrictions. The last three subsections explain the empirical puzzles from the introduction. My examples remain focused on IT outsourcing, but the multi-year contracts used in contract manufacturing,44 input procurement and franchising (see Brickley et al., 2003) also feature in the last three subsections. I refer to the Center for Organizations Research and Innovation (CORI) for detailed IT outsourcing contracts: Cobancorp-EDS (1995-2002), UHS-Unisys (1996-2005), GM-EDS (1996-2006), Tri City Bank-M&I Data Services (1998-2006).

investment (though trade can only occur once). They focus on the infinite-horizon setting.

42See Aghion and Bolton (1987) who distinguish this from the “actual” trade duration, of physical trading given renegotiation.

43For ever-green contracts, the length of the advance notice period is the key proxy, but the size of the penalty on giving notice is particularly important as the effective length is infinite if this penalty is too high.

44Witness the huge 3-year deals between IBM and Sanmina-SCI in 2002 and 2003, HP’s 5-year deal with Solectron in 2003, Casio’s 3-year deal with Flextronics in 2002, and Avaya’s 5-year deal with Celestica in 2001.
6.1 Contract Design

Three minor extensions of the basic model greatly increase the realism. First, traders generally choose among several possible basic trades when writing their initial performance contract. They optimally pick the trade contract for which valuable investments have high compatibility and undesirable investments have low compatibility.\(^{45}\) It is instructive to revisit Edlin and Reichelstein’s (1996) model in which specific performance contracts vary trade quantity to induce first-best investment incentives.\(^ {46}\) Introducing adaptation investments into their model would prevent the first-best, because raising trade quantity (or intensity) reduces side-compatibility (see 6.4 below). In my multi-period model, an extremely brief but intense contract could limit this market-shielding problem, but would probably fail to protect the right specific investments, instead encouraging wasteful investments that are only useful for producing at artificially high intensity.

Second, the long-term contract might include a trade menu from which at least one party (usually the client) can choose after investing. For instance, in IBM and Amex’s “pay-as-you-use” contract, Amex owes less to IBM if it uses less server space. This raises side-compatibility for any Amex adaptation that affects its server space needs.\(^ {47}\) However, flexibility must be limited to protect the seller’s specific investments. Pricing is generally non-linear (as in Goldberg and Erickson, 1987), and the lion’s share of Amex’s service charge is effectively fixed.\(^ {48}\)

Third, the contractual menu might be determined over time through a set of modification procedures. A large share of the vendor’s compensation, especially in the longer contracts used to support reorganization investments in “acquisition” deals like that of Amex and IBM,\(^ {49}\) often comes from a “revenue commitment” that fixes how much the client must spend on the vendor’s services (perhaps, as a percentage of the client’s IT expenditure - see GM-EDS, 1996). This is highly flexible, since trades and prices are determined by subsequent arbitration and benchmarking procedures. These procedures (complemented by reputational concerns) work well for minor ‘modifications’ and easily priced ‘new releases,’ but arbitration is imperfect and may specify an excessive price on an optional service.\(^ {50}\) Bilateral negotiations therefore play a

\(^{45}\)The statement applies with opposite signs for cross effects.

\(^{46}\)Their result holds for bilateral self-investments under a separability condition.

\(^{47}\)Recall Diversei’s investment in the adaptation of shifting to a client-server system which minimized its need for space on its outsourced mainframe server.

\(^{48}\)IBM’s claim to have invented a fully flexible “e-business utility” model is exaggerated (see Koch, 2003).

\(^{49}\)IT contracts for medium sized deals typically last between 3 and 5 years (as in Beulen and Ribbers, 2003).

\(^{50}\)Some contracts (see NAO, 2001, on Inland Revenue-Accenture) specify annual limits on “system enhancements” that the client can demand at a predetermined price. Flexibility is then huge but not when the limits are exceeded (as happened after legislators introduced “stakeholder pensions” in 1998 and quarterly tax reporting
key role and the credibility of market threats is critical.\textsuperscript{51}

\section*{6.2 Contract Enforcement}

I assumed complete non-verifiability of investment costs and payoffs, thereby ruling out reliance and expectation damages and restricting to liquidated damages. Relaxing this assumption reduces the difficulty of motivating investments, but unless measurement of costs or potential returns to adaptation is perfect, market forces and side-compatibility still matter. Note that I include here measurement by non-competing third parties, because the basic trade is partly enforced by ‘market reputation’ pressures to follow the spirit, not just the letter, of the bilateral contract. The relevant informational constraints for bilateral self-enforcement (trust, relational contracting) are usually weaker, but this topic lies beyond the scope of the paper.\textsuperscript{52}

The optimal design may require intermediate damages. In section 4, I explained how this enables option contracts to induce breach after low cross-investment. Intermediate damages also make breach contingent on the stochastic value of adaptation investments. For instance, after Halifax acquired the Bank of Scotland (BoS) in 2002, its value from substituting BoS’s centralized IT infrastructure with a more decentralized system justified paying the breach damages in BoS’s outsourcing contracts with IBM and Xansa (see Computer Weekly, 2002). This breach was endogenously contingent on the acquisition. Lowering breach damages raises side-compatibility, but also reduces the contract’s self-investment effect. Optimal design tries to induce credible breach in states where the marginal returns to adaptation investments are maximal, because this exploits market forces at their most powerful. Intermediate breach damages therefore dominate stochastic enforcement with high breach penalties, if the marginal and absolute returns on adaptations are positively correlated.\textsuperscript{53} Also, where feasible, breach damages are directly contingent on events, such as mergers and acquisitions, where adaptations tend to be important. This is why breach damages often fall gradually with time over the contract duration.

\textsuperscript{51}I define changes that need such negotiation or market threats to be adaptations. For example, when GM-EDS’s $40bn contract specifies “market tests” to benchmark EDS’s prices on additional services, these new services count as adaptations.

\textsuperscript{52}The implications for contract length are not obvious. Short-term contracting might support self-enforcement through termination threats, but being forced to trade together after trust has broken down may be a worse mutual punishment. Note also that formality need not imply distrust, and even the litigation between Xerox and EDS did not prevent them from later extending their trading relationship (DiSabatino, 2001). Baker et al’s (2002) repeated games extension of the formal work on property rights will offer an instructive parallel.

\textsuperscript{53}A mathematical example is available from the author.
6.3 Multiple Vendors

In 2003, Procter and Gamble (P&G) outsourced to three key vendors, HP, IBM and Jones, Lang & Lasalle - see Weiss and Perez (2003). A stylized analysis of this decision clarifies the flexibility advantage of the multi-vendor approach.\(^{54}\) Each vendor had to sink fixed costs in learning about P&G. Had P&G instead outsourced to EDS in a one-stop contract as initially planned, EDS would have reaped economies of scope by sinking one, instead of three sets of fixed learning costs. However, P&G benefits from the multiple sunk costs because these three vendors then create credible competition for many of the service adaptations that P&G will later discover. This competitive effect (increased side-compatibility\(^{55}\)) allows P&G to negotiate adaptations more cheaply and more reliably. In general, multi-sourcing implies high side-compatibility which raises the flexibility to adapt and permits longer contracts.

López and Ventura (2001) and González and López (2002) analyze outsourcing contracts in the electronics industry using primary source data from ANIEL\(^{56}\). They use a survey to measure how many subcontractors are used by each manufacturer and to find out the length of written contracts. Having multiple subcontractors, or vendors, is a good proxy for the existence of competing providers. So, based on Joskow (1987) (see also section 3 above), they expected a negative correlation with contract length. Instead, they found an insignificant positive correlation,\(^{57}\) but the puzzle is resolved by taking account of side-compatibility. Multi-sourcing reduces the cost of long-term contracting, because it increases side-compatibility. Proposition 5 shows that contract length is generally increasing in side-compatibility. This effect counteracts the pressure to make contracts shorter, and can therefore explain the empirical result.

The side-compatibility advantage of multi-sourcing can sometimes be achieved through “selective outsourcing,” whereby a client retains a strong internal IT unit. Amex chose this approach. Its internal IT unit credibly competes with IBM to satisfy some of its adaptation needs, so Amex’s side-compatibility is not so low.\(^{58}\) See Lacity et al. (1996) for evidence suggesting that selective outsourcing increases flexibility.

\(^{54}\)Experts emphasize the flexibility advantage. With multiple vendors, each vendor could be more specialized, but often they are not, so the specialization explanation is insufficient.

\(^{55}\)Fixed costs otherwise reduce the competition for side trades to a greater extent than they reduce competition for the larger trade that is up for bidding when long-term contracts expire.

\(^{56}\)Asociación Nacional de Industrias Electrónicas y de Telecomunicaciones.

\(^{57}\)Beulen and Ribbers’ (2003) IT outsourcing case studies exhibit the same phenomenon, but one should control for size which could proxy for reorganizations and the importance of specific investments.

\(^{58}\)Technically, the internal unit is part of Amex, but the general notion of competition remains valid.
6.4 Multiple Clients and Dedication

The case of coal processing is an empirical classic. Following Joskow (1987), Kerkvliet and Shogren (2001) measure the fraction of each coal company’s annual production capacity dedicated to a given contract for supplying coal to a utility. They were surprised to find a significant negative correlation between this measure of dedication and contract length. This finding is indeed anomalous given that high dedication plausibly proxies for relationship-specificity which is associated with longer contracts. The side-compatibility perspective offers two explanations. First, high dedication to a single contract may leave the vendor with insufficient capacity to sell adapted services to alternative clients while satisfying its contractual obligations. So side-compatibility is low and this induces shorter contracts (again, by proposition 5). Second, dedication also proxies for coal vendors with only one client. These vendors have low side-compatibility, because they have only paid one set of client-specific fixed costs.\textsuperscript{59} The anomaly is explained if either of these side-compatibility effects dominates the specificity effect.

6.5 Contractual Restrictions

In my basic model, long-term contracts impose trade obligations without shifting residual control rights, but rights and restrictions are sometimes needed to complement performance contracts. For instance, exclusive territory clauses complement franchise contracts by preventing franchisors from setting up new units that would compete nearby. As complementary tools for protecting specific investments by franchisees, exclusive territories and contract length should be correlated. Bercovitz (2000) tests for a positive correlation in her rich franchising data set. The puzzle is that she finds no significant correlation. The side-compatibility perspective offers a simple resolution: Exclusivity clauses reduce the side-compatibility of franchisor investments that require local adaptations, and this makes contract length more costly. If contracts do not always need exclusivity as a complement, then contracts will be longer in these cases (by proposition 5). This effect can explain the negative correlation.

She also finds an insignificant correlation between exclusivity and specificity. The absence of correlation may be caused by measurement difficulties. Franchisor estimates of franchisee set-up costs are used as a proxy for specific investments. However, this proxy (following Dnes, 1992) has been successful in other work, so I offer an alternative explanation based on side-compatibility: When specificity is low, exclusivity can be used to complement contracting, because the contract is then very short and each party can earn a reasonable return on its adaptation investments by waiting till the contract expires. Future empirical work that iden-

\textsuperscript{59}The argument is analogous to that for multi-sourcing.
tifies an exogenous source of variation in the use of exclusivity clauses would offer an excellent way to test this aspect of my theory more directly.\footnote{Variation in the legality of exclusivity restrictions across U.S. states presents a promising avenue, but restrictive states also tend to restrict contract termination. This introduces correlated bias in the use of reported contract length as a proxy for contract length, so careful measurement will be needed.}

Contracts regularly seek the opposite of exclusivity restrictions. Instead, they include rights to solicit competitive bids for services beyond the baseline contract.\footnote{This was common in Goldberg and Erickson’s (1987) data on coke supply contracts. It is also very common in IT contracting, but enforcement is difficult. For instance, Lacity and Wilcocks (1998) report two cases where vendors prevented third-party applications and hardware by claiming interference with their maintenance obligations.} Additionally, IT contracts and industry norms often require standardization and documentation of code and even pressure providers to cooperate,\footnote{E.g., clause 3.05 in UHS-Unisys obliges Unisys to support third-party software purchased by UHS.} greatly reducing transition costs (Beulen and Ribbers, 2003) and increasing side-compatibility. Further work should extend my theory to analyze how the allocation of rights can complement performance contracting in settings where ownership rights can be used to prevent third party involvement in making adaptations.

\section{Concluding Discussion}

I have analyzed the optimal duration of performance contracting in a bilateral model where both traders make multiple investments. My most important result is that contracts should be shorter when it is important to motivate adaptation investments, and when side-compatibility is low. Side-compatibility is high when a trader has multiple vendors, when long-term contracts do not exhaust the capacity or need for trade, and when there are no exclusivity clauses restricting side-trades. In each case, my theory predicts that high side-compatibility will pressure towards longer contracts. Using the three cases, I resolved three empirical puzzles from the transaction cost literature.

My market-shielding perspective endogenizes and qualifies the transaction cost idea that long-term contracts have flexibility costs. Formally, I have shown that adaptations are cheaper under short-term contracting, unless side-compatibility is complete. My framework can also show why renegotiating from a long-term contract is likely to be harder than negotiating from scratch:~\footnote{Many theorists have assumed there is no difference, positing instead an exogenous cost of renegotiation (see e.g., Harris and Holmström, 1987, and Guriev and Kvasov, 2004).} Unless side-compatibility is complete, long-term contracts reduce competition for adaptations. Since competitive bids tend to reveal information about the cost of adaptations, thereby reducing information asymmetries, this competition reduces the risk of negotiation.

60 Variation in the legality of exclusivity restrictions across U.S. states presents a promising avenue, but restrictive states also tend to restrict contract termination. This introduces correlated bias in the use of reported contract length as a proxy for contract length, so careful measurement will be needed.

61 This was common in Goldberg and Erickson’s (1987) data on coke supply contracts. It is also very common in IT contracting, but enforcement is difficult. For instance, Lacity and Wilcocks (1998) report two cases where vendors prevented third-party applications and hardware by claiming interference with their maintenance obligations.

62 E.g., clause 3.05 in UHS-Unisys obliges Unisys to support third-party software purchased by UHS.

63 Many theorists have assumed there is no difference, positing instead an exogenous cost of renegotiation (see e.g., Harris and Holmström, 1987, and Guriev and Kvasov, 2004).
failure. So long-term contracts increase the risk of failing to negotiate efficient adaptations.\textsuperscript{64} My results formally corroborate prior insights from transaction cost theory on the effects of market-thickness, investment specificity and uncertainty. The model is very tractable. Cross effects and wasteful investments have the exact inverse impact of the corresponding self effects and non-wasteful investments and I readily analyzed temporal variation in investment productivity. By explicitly evaluating the tradeoffs between multiple investments, I identified how level effects sometimes counteract the intuitive substitution effects on optimal length. I also showed how intermediate breach penalties and menu contracts can raise contractual flexibility. Relational contracting may further enhance the effectiveness of menu contracts, so this presents important work for future analysis. Extending the model to allow for partially measurable costs is also interesting: Bajari and Tadelis (2001) argue that cost-plus contracts are more flexible than fixed-price contracts because renegotiation is then more efficient. They do not consider market forces, so combining our models could generate novel predictions. For instance, I conjecture that settings with high side-compatibility will favor fixed-price contracting by mitigating the flexibility problem.

Because of a false presumption that general investments are not at risk of holdup, other formal analyses have neglected how long-term contracts interfere with valuable market forces. Most papers with long-term contracting simply ignored general investments. MacLeod and Malcomson (1993) is an exception, but they only allowed for one joint trade, thereby precluding the possibility of general investments that require an adapted joint trade. The closest analogy to my market-shielding result is in fact Hart and Moore’s (1990) original analysis of ownership. In a setting without long-term contracting, they showed how removing ownership rights interferes with access to market forces that could otherwise motivate asset-specific investments. In a setting with long-term contracts, I have shown how a contractual obligation to trade shields out market forces that could otherwise motivate adaptation investments. My model of this simple point has already helped resolve three empirical puzzles. Collecting empirical data with an eye to creating effective proxies of side-compatibility will permit tighter tests of my predictions, and will, I believe, improve our ability to explain contract duration.

\textsuperscript{64}Future work that extends my model to include alternative traders that invest and bid to win new trades, could analyze more precisely how flexibility rises with side-compatibility.
Appendix

**Proof of Proposition 1.** Set \( \alpha = 1 \) in Proposition 3. \[\square\]

**Proof of Proposition 2.** Set \( \alpha = 0 \) in Proposition 3. \[\square\]

**Proof of Proposition 3.** Differentiating the identities in (2) with respect to \( \alpha \) gives,

\[
\frac{d e_j}{d \alpha} = \left[ -W_j' (e_j) \right] \frac{\psi_{e_j} - \gamma_{e_j}}{1 + \alpha \cdot \psi_{e_j} + (1 - \alpha) \gamma_{e_j}} \quad (A1)
\]

\[
\frac{d i_j}{d \alpha} = \left[ -V_j' (i_j) \right] \frac{\psi_{i_j} - \gamma_{i_j}}{1 + \alpha \cdot \psi_{i_j} + (1 - \alpha) \gamma_{i_j}}
\]

The bracketed expressions are positive by Assumption 1. \( F = 1 + \alpha \cdot \psi + (1 - \alpha) \gamma \) is always positive by Assumption 2. The sign of \( \psi - \gamma \) is positive for \( e_P \) and \( e_A \) and negative for \( i_P \) and \( i_A \), by type 1 and 2 definitions. This proves 3(a).

Differentiating with respect to \( \gamma \) and \( \psi \) gives,

\[
\frac{d e_j}{d \gamma_{e_j}} = \left[ -W_j' (e_j) \right] \frac{1 - \alpha}{1 + \alpha \cdot \psi_{e_j} + (1 - \alpha) \gamma_{e_j}}
\]

\[
\frac{d e_j}{d \psi_{e_j}} = \left[ -W_j' (e_j) \right] \frac{\alpha}{1 + \alpha \cdot \psi_{e_j} + (1 - \alpha) \gamma_{e_j}}
\]

Both these expressions are positive for \( \alpha \in (0, 1) \). The first is positive even when \( \alpha = 0 \), the second is positive even when \( \alpha = 1 \). This proves 3(b) since the exact same expressions and observations are valid for \( i_j \). 3(c) follows from the observations that \( e_j (1) = e_j^{LTC} \leq e_j^* \) and \( i_j (0) = i_j^{STC} \leq i_j^* \) by comparing the first order conditions (2) with \( \alpha = 1 \) and 0, with the conditions (1) for the first-best. \[\square\]

**Proof of Proposition 4.**

I write the first-order condition for (5) as \( D = 0 \) where,

\[
D \equiv \sum_{j=P,A} \left( E_j W_j' (e_j (\alpha, E_j)) - 1 \right) \frac{d e_j (\alpha, E_j)}{d \alpha} + \left( I_j V_j' (i_j (\alpha, I_j)) - 1 \right) \frac{d i_j (\alpha, I_j)}{d \alpha}
\]

\[
= \sum_{j=P,A} \left( D_{e_j} + D_{i_j} \right)
\]

Applying the Implicit Function Theorem to this first-order condition \((D (E, I; \alpha (E, I))) \equiv 0\) gives,

\[
\frac{d \alpha}{d E_j} = -\frac{\frac{\partial D}{\partial E_j}}{\frac{\partial D}{\partial \alpha}} \quad (A2)
\]

Assumption 4 ensures that \( \frac{\partial D}{\partial \alpha} < 0 \) - see below for a sufficient condition. Defining \( \delta_{e_j} \equiv \psi_{e_j} - \gamma_{e_j} \), the rescaled version of (A1), states that \( \frac{d e_j}{d \alpha} = \left[ -W_j' (e_j) \right] \frac{\delta_{e_j}}{W_j' (e_j)} = \frac{-\delta_{e_j}}{W_j' (e_j)} \).
Substituting this and first-order condition (5) into $D_{e_j}$, and differentiating $D$ with respect to $E_j$ gives,

$$
\frac{\partial D}{\partial E_j} = \frac{\partial D_{e_j}}{\partial E_j} = \frac{\partial}{\partial E_j} \left( \left( \frac{2}{F} - 1 \right) E_j W_j''(e_j(\alpha, E_j)) \frac{2}{F^2} \right) = \frac{2}{F^2} \left( \frac{2}{F} - 1 \right) \delta_{e_j} \frac{\partial}{\partial E_j} \left( E_j W_j''(e_j(\alpha, E_j)) \right)
$$

As $F < 2$ by assumption 1 and $\delta_{e_j} > 0$ by definition of type 1, this has the same sign as $\frac{\partial}{\partial e_j} \left( E_j W_j''(e_j(\alpha, E_j)) \right)$. Differentiating (5) $(\frac{2}{F}F(\alpha; \psi_{e_j}, \gamma_{e_j})E_j \cdot W_j'(e_j(\alpha, E_j)) \equiv 1)$ with respect to $E_j$ and substituting, shows that this equals,

$$W_j'' + E_j W_j''' \frac{\partial e_j}{\partial E_j} = W_j'' + E_j W_j''' \frac{-W_j'}{E_j W_j''}
$$

As $W_j'' < 0$ by assumption 1, this has the same sign as $W_j' W_j''' - (W_j'')^2$ which is positive by assumption 3. Hence, $\frac{\partial D}{\partial E_j} > 0$ so (A2) reveals that $\frac{\partial}{\partial e_j} > 0$ as claimed.

**Conditions for Assumption 4:** A sufficient condition is $W_j'' W_j' < 4 \left( W_j'' \right)^2$ and $V_j'' V_j' < 4 \left( V_j'' \right)^2$. Note that this condition is satisfied by common functions that also satisfying assumptions 1 and 3 - e.g., whenever $W(\cdot)$ and $V(\cdot)$ take any of the forms: $\sqrt{e}$, $-\frac{1}{e}$ and $\log e$.

**Proof of sufficiency.**

$$
\frac{dD_{e_j}}{d\alpha} = \frac{d}{d\alpha} \left( (W_j' - 1) \frac{de_j}{d\alpha} \right) = W_j'' \left( \frac{de_j}{d\alpha} \right)^2 + \left( W_j' - 1 \right) \frac{d^2 e_j}{d\alpha^2} = \frac{1}{W_j'} \left[ \left( W_j'' \frac{de_j}{d\alpha} \right)^2 + \left( W_j' - 1 \right) \left( \frac{4\delta^2}{F^3} - W_j'' \frac{de_j}{d\alpha} \right)^2 \right] = \frac{1}{W_j'} \left[ \frac{4\delta^2}{F^4} \left( 3 - F \right) - 2 - F \frac{W_j''}{W_j''} \right] = \frac{1}{W_j'} \left[ \frac{2\delta^2}{F^4} \left( 6 - 2F \right) - (2 - F) \frac{W_j' W_j''}{(W_j'')^2} \right] \equiv W'' W' - \left( 2 + \frac{2}{2 - F} \right) (W'')^2
$$

and $2 + \frac{2}{2 - F} \geq 4$ so this is negative; and $\frac{d\left(-D_{e_j}\right)}{d\alpha} \equiv V'' V' - \left( 2 + \frac{2}{2 - F} \right) (V'')^2$ takes the opposite sign so $\frac{dD}{d\alpha} < 0$.

**Proof of Proposition 5.** I prove for the case of $e_j$ (same arguments hold for $i_j$). To study changes in $\delta$ with $F$ fixed at $\bar{F}$ for $\alpha = \bar{\alpha}$, differentiating $D$ at $\alpha = \bar{\alpha}$ simply gives:

$$
\frac{\partial D}{\partial e_j} \big|_{F_{e_j}(\bar{\alpha}) = \bar{F}_{e_j}} = \left( \frac{2}{F_{e_j}} - 1 \right) \frac{2}{F_{e_j}^2} \left( -\frac{1}{W_{e_j}} \right) > 0
$$

(as $\bar{F}_{e_j} < 2$ and $W'' < 0$). Applying the
Implicit Function Theorem (IFT) to the first-order condition \( (D(E, I, \psi, \gamma; \alpha (E, I; \psi, \gamma)) \equiv 0) \) gives,

\[
\frac{\partial \alpha}{\partial \delta} \bigg|_{F_{e_j}(\hat{\alpha}) = F_{e_j}} = -\frac{\frac{\partial D}{\partial \delta_{e_j}} \bigg|_{F_{e_j}(\hat{\alpha}) = F_{e_j}}}{\frac{\partial D}{\partial \alpha}}
\]

Since \( \frac{\partial D}{\partial \alpha} < 0 \), this substitution effect is positive at \( \alpha = \hat{\alpha} \), so \( \alpha \) increases as claimed.

Similarly, the IFT gives the level effect of changing \( F_{e_j} \) with \( \delta_{e_j} \) fixed at \( \delta \): \( \frac{\partial \alpha}{\partial F_{e_j}} \bigg|_{\psi_{e_j} - \gamma_{e_j} = \delta} \) sign

\[
= \frac{1}{\beta_{e_j}} \frac{\partial D_{e_j}}{\partial \alpha} \bigg|_{\delta_{e_j}}
\]

since \( F_{e_j} \equiv \left(1 + \gamma_{e_j} + \alpha \delta_{e_j}\right) \) and \( D_{e_j} \) depends on \( \alpha \) through \( F_{e_j} \). Now \( \frac{\partial D_{e_j}}{\partial \alpha} < 0 \), so this level effect has the same sign as \(-\delta_{e_j}\), exactly as claimed.

References


[16] CORI, Contracting and Organizations Research Institute, University of Missouri-Columbia (http://cori.missouri.edu)


34