On the Grouping of Tasks into Firms:

Make-or-Buy with Interdependent Parts*

by

Sharon Novak
snovak@mit.edu

and

Birger Wernerfelt
Massachusetts Institute of Technology
Sloan School of Management
bwerner@mit.edu

December 2, 2008.

JEL Codes: D23, L22, L62

Keywords: Division of labor, Make-or-buy, Adjustment-costs, Automobiles

*An early version of this paper was entitled "The Design of Industry". We are indebted to seminar participants at Berkeley, Harvard, MIT, NYU, Toronto, the 2006 IIOC, and the 2007 AEA meetings, as well as George Baker, Ken Corts, Mercedes Delgado Garcia, Bob Gibbons, Oliver Hart, Tom Hubbard, Whitney Newey, Jim Orlin, Roberto Rigobon, Duncan Simester, Eric Van den Steen, Tom Stoker, Dan Whitney, Dennis Yao, Florian Zettelmeyer, and Ezra Zuckerman for helpful comments, and to Dmitri Byzalov, Chris Ruggerio, and especially Carol Meyers for excellent research assistance. Of course, the usual disclaimer applies.

On the Grouping of Tasks into Firms:

Make-or-Buy with Interdependent Parts*

Abstract

We study the division of labor within a supply chain and look for the optimal grouping of tasks into firms. Using a unique dataset on supply chains in the global automobile industry, we present robust evidence consistent with the view that the firm is a low variable-, but high fixed cost way to govern adjustments. The results are strong, suggesting that the theory captures an important effect. By taking account of interdependencies between parts, our econometric approach generalizes standard make-or-buy analysis and yields improvements in predictive accuracy.

What determines how the sub-tasks of a production process are grouped into firms? This is a slightly novel way of asking a question that is central to our understanding of the nature of coordination within and between firms. Many classical theories in the field suggest that the inefficiencies of markets increase with the number of transactions, while the costs of integration are more or less fixed.[1] We here examine a similar, but subtly different proposition: that the grouping of tasks is influenced by the frequency with which they have to be coordinated. Using unique data on eight supply chains in the global automobile industry, we test the proposition that the firm is a low variable-, but high fixed cost way to govern adjustments.

Because our hypothesis involves the entire supply chain, the empirical specification relaxes an independence assumption implicitly made in most studies of make-or-buy decisions.[2] From the perspective of our model, a make-or-buy study starts with a focal task and works off the relationship between that and each of the other tasks, while ignoring relationships between pairs of non-focal tasks. For example, suppose that tasks 1 and 2 should be together and that tasks 2 and 3 should be together as well. In this case tasks 1 and 3 will be together independently of their direct relationship, and it is meaningless to perform a make-or-buy analysis of the question.

[1] In particular, Coase (1937) emphasized the need to avoid the haggling costs associated with market transactions, Simon (1951) used implicit super game arguments to suggest that more frequent trades should be internalized, and Williamson (1979) similarly argued that higher frequency of trade, combined with asset specificity, favors integration.

[2] For example, Monteverde and Teece (1982), Novak and Eppinger (2001), Simester and Knez (2002), and Nagaoka, Takeishi, and Noro (2008). Lafontaine and Slade (2007) present a quite complete review of this very large literature.

In a model that takes all pair-wise relationships are taken into account simultaneously, we show: (1) that any two tasks are more likely to be performed by the same firm if mutual adjustments are needed on a sufficiently frequent basis, (2) that a disproportionate share of all adjustments are managed inside firms, and (3) that supply chain design can be portrayed as the solution to an integer program aimed at minimizing the sum of adjustment-costs within and between firms.[3] The results are strong and robust, suggesting that the theory captures an important effect. To evaluate the practical importance of the above-mentioned independence assumption, we compare the analysis with a standard make-or-buy estimation and find that the predictive accuracy is much improved by taking into account the interdependencies between tasks.

While our test is more powerful, data requirements and computational difficulties are obstacles to its use. For each pair of parts, we need data on the frequency with which mutual adjustments are needed; and for each part, we need to know its producer. Publicly available sources do not contain this type of information, and we had to make an extremely large investment in data collection in order to acquire it. The second problem concerns the estimation procedure. Even for small sets of parts, there are a very, very large number of possible groupings of tasks, each of which needs to be considered as a counterfactual.

---

[3] While few economists have studied supply chains, they have received a lot of attention in operations management, primarily in the context of the "Design Structure Matrix" (Steward, 1981; Eppinger, 1991; and Baldwin and Clark, 2000). This matrix summarizes the direction and importance of information flows between pairs of tasks, and is a tool for managing new product development processes. It contains more and different information than that used here and the question of firm boundaries is of little interest to this literature. However, it is interesting to note that Baldwin and Clark (2000, p. 368ff.) adopt arguments similar to those tested here and suggest that firm coordination carries low variable costs compared to market coordination.

Since the dependencies cause extreme non-linearities, we use simulated GMM to estimate the model. A very large integer programming problem has to be solved in each simulation, and we believe that our 36 parts are close to the limit of what is currently practical. On the other hand, the efforts at data collection and computation have given us a very well-fitting structural model.

After a brief review of the theory, we derive the hypotheses in Section II. The data are described in Section II, and estimation techniques and results are presented in Section III. In Section IV, we briefly compare our estimates with those from a traditional make-or-buy model, and the paper concludes with a discussion in Section V.

I. Theory and Hypotheses

In spite of its central importance to the field, economists have not yet agreed on a theory of the firm. The several competing theories include, but are not limited to, those of Coase (1937), Grossman and Hart (1986), Hart and Moore (2008), Holmstrom and Milgrom (1994), Simon (1951), and Williamson (1979). A significant amount of empirical work has failed to settle the issue, possibly because many of the theories are difficult to conclusively falsify with the datasets used.

In the present paper, we are able to perform a direct test of the adjustment-cost theory of the firm (Wernerfelt, 1997). There are two reasons for this. First, our data map precisely to the central prediction of the theory - that tasks in need of more frequent mutual adjustments are grouped together. Secondly, the prediction is robust across a variety of extensive forms.

Given this, we now summarize the underlying argument. Consistent with Hart's (2008) call for theories relying on non-Coasian bargaining, the adjustment-cost theory of the firm compares alternative game forms in light of the costs of change (adjustment). In its simplest form, it looks at a trading relationship between two players in which maximization of joint payoffs occasionally requires that the actions of one or both players change. The theory looks for the most efficient game form with which the players can agree on the adjustments and associated changes in payments. There are a sea of possibilities, but for the present purposes we will compare two alternatives: "Negotiation-as-needed" in which the players have a bilateral discussion about each adjustment and associated changes, and an "Employment Relationship" in which one player ex ante agrees to let the other dictate what happens in case of adjustment, while both retain the right to terminate the relationship.[4] Purchases governed through these two game forms could be a house renovation and secretarial services. The theory is controversial because the adjustment-costs, through which it compares the game forms, essentially are bargaining costs.[5]

Costs of bargaining may be incurred before, during, or after the process.[6] The most obvious source is the time consumed during bargaining. Any explicit model of alternating offer bargaining must posit some costs of refusing an offer

[4] Since players can save pricing costs in a wide set of circumstances, it is not uncommon to see hybrid game forms with employment-like elements in other game forms. For example, a haircut can typically be "as you like it". However, consistent with everyday language, the term "employment relationship" is reserved for the extreme case in which one player's discretionary power includes at least some adjustments to the work methods used by the other. (You could not ask the beautician to cut with the other hand.)

[5] In the simplest versions, we assume that all efficient adjustments are implemented, such that gains from trade have no impact on the comparisons.

[6] The exact nature of the bargaining costs is not essential to the argument.

and making a counter, since the process otherwise would go on ad infinitum. Delays are strictly out-of-equilibrium outcomes in the most simple models (Rubinstein, 1982), but not in richer settings (Watson, 1998). Consistent with this, a survey by *Purchasing* Magazine suggests that US purchasing managers spend 15% of their time on price negotiations. Costs incurred prior to bargaining may be an even more important source. It is well-documented that better-informed bargainers get better results (Busse, Silva-Risso, and Zettelmeyer, 2006), and while this result does not figure prominently in the theoretical literature, it is easy to understand. The idea is that players, prior to bargaining, can invest to get information that will improve their shares, but not the overall gains from trade. Consistent with the importance of anticipatory bargaining costs, the above-mentioned survey of purchasing managers found that they spent about 25% of their time "Preparing Bids" and "Researching Prices". Finally, Hart and Moore (2008) have recently argued that any not-ex-ante-agreed-upon outcome produces post bargaining aggrievement towards the trading partner and a reduction in gains from trade.

Given the existence of bargaining costs, the argument is simple. Since the employment relationship involves less negotiation per adjustment, it has lower variable costs. On the other hand, employment entails some fixed costs ex ante because the parties have to reach agreement on the arrangement. There are no such fixed costs with negotiation-as-needed because no ex ante agreement is needed to open negotiation ex post. So among these two game forms, "Negotiation-as-needed" is most efficient when adjustments are rare and the

"Employment Relationship" is the lowest cost solution when adjustments are both frequent. Different implications of the theory have been tested before, for example by Simester and Knez (2002), but the present paper is the most extensive test by far.

The product consists of N parts, and for each of the $N(N-1)/2$ pairs i and j we know the frequency with which there has to be a coordinated adjustment in them.[7] We denote this variable by a_{ij}. Since the parts are complex in themselves, there is no automatic transitivity in the need for adjustment. That is, it is possible that $a_{12} > 0$, $a_{23} > 0$, and $a_{13} = 0$. The first two sets of adjustments could be necessary without the third being so (as if the first is about color and the second is about size).

The set of parts $\mathbf{N}$ ($|\mathbf{N}|=N$) can be made by anywhere from one to N firms and we define a supply chain design (aka a grouping of tasks) as a partition $\mathbf{S} = \{\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_n}\}$ of $\mathbf{N}$ into $n \leq N$ non-empty, non-overlapping sets (aka firms). We take the frequency of needs for adjustments as exogenously given and assume that the parties can contract on designs, but not on ex post bargaining behavior.[8] Given this, we make the standard assumption that the players write the most efficient contracts and look for the design with the smallest total adjustment-costs.[9] While this is a choice, it is not unreasonable. Several supply chains compete in our empirical setting and it is fair to conjecture that the more

[7] For the present purposes, we only consider mutual adjustments betweens pairs of parts. We could more generally think of all higher orders up to N.

[8] While we thus treat technology, represented by the frequency of needs for mutual adjustments, as exogenously given, it is obviously endogenous in the very long run. For example, it takes up to ten years to launch a new technological platform in this industry

[9] Many papers on the theory of the firm, including all those cited at the start of this Section, make this assumption.

efficiently organized chains survive. The assumption is nevertheless strong and it would be nice to develop a non-cooperative foundation for it.[10] At the same time, the extreme complexity of the issues studied means that any justification in terms of equilibrium, and even survival, stretch common sense.

We use i and j as generic parts produced in firms I and J, respectively, while b and c are generic firms. The variable adjustment-costs, per adjustment per pair, are m in the market and $f(S_b) < m$ in a firm of size $S_b \equiv |\mathbf{S_b}|$, where the f's are non-decreasing. Total fixed adjustment-costs are $w(S_b-1)$ in a firm of size S_b (and 0 in the market). To keep things simple, we assume that all adjustments are implemented, regardless of the costs thereof. Recalling that I is the set of parts made by the same firm as part i, total expected costs are therefore $(N-n)w + \frac{1}{2}\Sigma_{i\in\mathbf{N}}[f(S_I)\Sigma_{j\in I}\, a_{ij} + m\Sigma_{j\notin I}\, a_{ij}]$, and the cost-minimizing grouping solves the Partitioning Problem (P):

$$\text{Min}_n\text{Min}_{\mathbf{S}}\ (N-n)w + \tfrac{1}{2}\Sigma_{i\in\mathbf{N}}[f(S_I)\Sigma_{j\in I}\, a_{ij} + m\Sigma_{j\notin I}\, a_{ij}], \text{ s.t.}$$

$$\bigcup_{b=1}^{n} \mathbf{S}_b = \mathbf{N},\ \bigcap_{b,c\neq b} \mathbf{S}_b\ \mathbf{S}_c = \emptyset, \text{ and } \mathbf{S}_b \neq \emptyset \text{ for all } b. \qquad (P)$$

There are a finite number of possible groupings, so while it is hard to characterize solutions, they trivially exist.

We test the theory from three different angles. We start by checking the first-order implication that two parts i and j are more likely to be made by the same firm if a_{ij} is higher. To account for indirect effects, we then look at the entire supply chain and find that a disproportionate number of adjustments are managed

[10] The most natural way to do this would be to look at each part as a player. However, since the industry originally was more, rather than less, integrated, this might not be the correct specification. In any event, the game is sure to be poorly behaved.

inside firms. We finally estimate a structural model in which supply chain design is found as a solution to a program like *P* aimed at minimizing total adjustment-costs.

II. Data

We have data from the late 1980s on eight supply chains pertaining to eight different cars in the luxury-performance segment of the automobile market. We look at each supply chain as a separate data set and will henceforth use that term and label them, or the cars they produce, A, B, C, D, E, F, G, and H. Since there is virtually no overlap between the participants in the eight supply chains, we treat them as independent, but identical.[11] That is, we assume that the same technology and the same set of parts, and therefore the same adjustment frequencies, drive the design of all eight supply chains. So we can estimate on a per-car basis, or pool the data across all eight.

A car consists of more than ten thousand parts.[12] Since we need information on *N(N-1)/2* pairs of parts, computational constraints and problems associated with administering very long questionnaires force us to divide the set of parts into 36 "megaparts", henceforth "parts", for the purposes of the study.

[11] Across the 36x8 part-car combinations, very few were subject to inter-car linkages: one firm made a part for 4 cars, three made a part for 3 cars, and eight made a part for 2 cars. In each of these cases, the firm was required to have separate employees and often even separate production facilities to serve each customer.

[12] A possible critique of any study of integration, and thus this, is that one could bias the results by including "irrelevant" parts. The idea is best brought out by fixing ideas on an example, such as office supplies. Suppose that some unknown factor causes office supplies to be outsourced and have zero mutual adjustment frequencies with all the other parts. In this case, the estimated relationship between in-sourcing and mutual adjustment frequencies would be stronger if office supplies were included in the sample. We have four responses to this critique. First, the exact same argument could be made about a traditional make-or-buy study testing the importance of specific investments. Second, all of our parts are components of the cars in question, and no components have been left out. Third, not a single one of our parts are produced in isolation across all eight supply chains. Finally, it is not clear what this "unknown factor" could be.

This aggregation is not arbitrary, but reflects industry use. More importantly, our data are such that all megaparts in all cars are assembled by a single firm. That is, there are no instances in which a firm delivers a fraction of one megapart plus another megapart. So no inconsistency is introduced by conducting the study at the megapart level, although we do lose a significant amount of information. In particular, we can not make use of the fact that the sub-components of the parts are co-produced, even though our theory speaks directly to it.

Since there are 36 “parts”, we have 630 pairs of parts, and we know the frequency with which mutual adjustment is needed as well as whether or not they are co-produced (made by the same firm)The (mega)parts are listed in Table 1 below.

Table 1

List of Parts

Body-in-white	Airbag controller	Intake manifold	Alternator
Body sheet metal	Airbag	Crankshaft	Speed control
Headlining	Power steering gear	Camshaft	Automatic transmission
Bumpers	Steering linkage	Piston	Suspension
Safety belts	Steering column	Intake valves	Drive shaft
Lock cylinders	Steering wheel	Radiator	ABS system
Door handles	Power steering pump	Starter	Spindle assembly
Windshield washers	Cylinder head	Distributor	Upper and lower arms
Seat system	Engine block	Instrument panel	AC assembly

The data on supply chain design (patterns of co-production), as well as some of our information about adjustment frequencies, come from interviews conducted by one of the authors. These interviews were very extensive and wide-ranging – involving more than 1000 employees of the eight firms. They included specific questions identifying the producer of every part as well as the need to coordinate between parts. Subjects were not systematically asked about adjustment frequencies, but the topic was repeatedly touched upon and the interviewer used her notes to construct estimates for all (part, part) pairs. Specifically, for each pair of parts, she rated "the frequency with which there needs to be a mutual adjustment in this pair" on a seven point scale from 0 - 6.[13] Most of the mutual adjustments between parts occur during the design phase (at the time of the interviews all parts were designed by the party producing them.): Several parts are strongly interdependent. For example the engine and the body, the pistons and the intake valves, the lock cylinders and the door handles, etc. Even small changes in one of these are likely to have implications for the other. Other pairs are less tightly linked, and most are for all practical purposes independent. [14]

A possible problem is that the interviewer rated the adjustment frequencies with knowledge of the hypothesis to be tested. As an ex post check, we therefore used a questionnaire to collect a second set of adjustment frequencies from an industry expert who was unaware of the hypothesis. Our expert is Dan Whitney, who for many years has played a major role in MITs International Motor Vehicle Program. After getting a table with the 630 (=36 x 35/2) pairs of parts, this expert

[13] Virtually all economic research based on questionnaire data use such Likert-type scales and proceeds to treat them as continuous in the statistical analysis. We follow this practice but realize that one can construct examples in which the ordinal scale corrupts the statistical inferences.

[14] Table A-1 in the Appendix displays the interview data on adjustment frequencies along with the co-production data for car A.

was asked to think of a typical luxury-performance car in the late 1980s, and answer the question: "Please consider a pair of parts and rate, on a scale from 0-6, the frequency with which there needs to be a mutual adjustment in this pair." It turns out that the questionnaire ratings are extremely similar to the interview ratings. The expert could have calibrated the seven point scale differently than the interviewer, but at .88 and .87 the means are almost identical, as are the fraction of frequencies rated zero (.72 and .74). Most importantly, there is a highly significant Spearman rank-order correlation of .915 between the two data sets. To keep the argument as clean as possible, all results reported in the body of the paper are based on the expert's questionnaire responses. Analog analyses, based on our interviews, are reported in the Appendix. As can be seen there, the results are essentially the same for both measures.

Some descriptive statistics are given in Table 2, in which the second column indicates the fraction of each car's parts that are co-produced, the third gives the sizes of all clusters of co-produced parts, and the *r'*s in the fourth column are Spearman rank-order correlations.

Table 2

Descriptive Statistics by Supply Chain[1]

Car	Fraction co-produced	Size-distribution	*r*(co-production, adjustment frequency)
A	.29	19, 5, 2, 2, 2	.53
B	.26	16, 9, 4, 2	.51
C	.14	13, 5, 2, 2, 2	.61
D	.18	15, 3, 3, 2, 2, 2, 2	.58
E	.79	32	.51
F	.17	14, 5, 3, 3, 2, 2	.54
G	.12	12, 4, 3, 2, 2, 2	.63
H	.11	11, 4, 3, 3, 2	.64

[1]N=630.

Looking first at the fraction of parts that are co-produced, we notice that all the supply chains are designed differently, with E as an extreme outlier.[15] This heterogeneity ought to make us pessimistic about our tests, since we hypothesize that all supply chains are solutions to the same optimization problem, but observe apparently very different designs. As a first clue that this pessimism would be unfounded, we see in all eight cases a very strong rank-order correlation between co-production and adjustment frequency. In spite of the differences, all the designs reflect a strong influence of adjustment frequency. [16]

[15] E represents a corporate form not used in the U.S., and it is possible to argue that the 32 parts are produced by three, rather than one, firm. However, we wanted to be as conservative as possible. It should also be noted that three of the other supply chains are anchored in the same country as E.

[16] If all observations were independent, the Spearman correlations in Table 2 would have *t*-values (~25*r*) above 10. However, as noted earlier, the data exhibits complicated dependencies, because "parts 1 and 2 are co-produced" and "parts 2 and 3 are co-produced" imply that "parts 1 and 3 are co-produced". So the correlations in Table 2 do in some sense overstate the degrees of freedom but are suggestive.

III. Estimation Techniques and Results

In this Section we present three successively deeper tests of the relationship between adjustment frequency and co-production. To assess the robustness of the theory, we have done most of these tests on a car-by-car, rather than pooled, basis. We start by looking at direct effects only, asking whether two parts are more likely to be co-produced if mutual adjustments are needed more frequently. This is done for each pair, while incorporating corrections for the interdependencies between the pairs. To take account of indirect effects, we then go to the supply chain-level and compare the sum of internalized adjustment ratings against the distribution of the same measure in random supply chain designs. We finally estimate a model in which supply chain design is portrayed as the outcome of a maximization problem aimed at internalizing a weighted average of adjustments and random noise. This allows us to evaluate the importance of adjustment costs relative to other forces in supply chain design.

III.i. Tests at the pair-by-pair level

There are many ways to draw statistical inferences about the relationship between co-production and adjustment frequency at the pair-by-pair level. We have chosen to compare the actual design (the actual pattern of co-production) to those in randomly generated allocations of the 36 parts to firms of the same size.[17] To this end we drew 100,000 randomly designed supply chains each having the

[17] It may seem more natural to compare against "completely" random designs generated by allocating each of the 36 parts to 36 equiprobable firms without the constraint on firm sizes. The problem is that all eight supply chains have quite skewed size-distributions, with the owner of the brand name being very large. Such large firms, and thus instances of co-production, would be relatively rare in a set of completely random designs, implying that our results would be much stronger. However, we do not feel comfortable attributing the skewed size-distributions to our theory alone (the brand name manufacturers are becoming less and less integrated, but are still making a very large share of all parts).

same size distribution as the actual. We then compared random and actual designs in terms on the probability that pairs of parts with identical adjustment frequencies are co-produced. The results are given in Table 3 below.

Table 3

Probability of Pairwise Co-production by Adjustment Frequency[1]

Car	0	1	2	3	4	5	6
A	.232 *(.001)*	.364 *(.234)*	.500 *(.089)*	.517 *(.020)*	.286 *(.598)*	.500 *(.012)*	.867 *(.001)*
B	.221 *(.008)*	.273 *(.474)*	.250 *(.605)*	.448 *(.026)*	.314 *(.272)*	.406 *(.050)*	.600 *(.019)*
C	.105 *(.002)*	.242 *(.139)*	.150 (.533)	.379 *(.004)*	.257 *(.056)*	.188 *(.306)*	.333 *(.095)*
D	.141 *(.006)*	.152 *(.681)*	.200 *(.491)*	.345 *(.045)*	.257 *(.173)*	.313 *(.057)*	.733 *(.000)*
E	.761 *(.087)*	.939 *(.113)*	.800 *(.662)*	.897 *(.164)*	.771 *(.683)*	.844 *(.301)*	.933 *(.283)*
F	.150 *(.047)*	.121 *(.795)*	.150 *(.650)*	.345 *(.029)*	.229 *(.245)*	.281 *(.089)*	.333 *(.140)*
G	.090 *(.005)*	.091 *(.745)*	.050 *(.887)*	.241 *(.070)*	.229 *(.059)*	.250 *(.038)*	.667 *(.001)*
H	.077 *(.004)*	.030 *(.963)*	.200 *(.194)*	.172 *(.202)*	.257 *(.007)*	.125 *(.458)*	.600 *(.000)*

[1] *p*-values in parentheses refer to tests relative to 100,000 randomly designed supply chains.

Since there is a lot of information in Table 3, we will briefly look at the interpretation of a couple of cells. The .232 in the "A, 0" cell means that 23.2% of all pairs rated 0 were co-produced in supply chain A. Recalling from Table 2 that 29% of all pairs are co-produced in this supply chain, a random design would have that percentage instead of the 23.2 and the statistical significance (.001) refers to a test versus that benchmark. Similarly, the .333 in the "C, 6" cell means that 33.3% of all pairs rated 6 were co-produced in supply chain C. Given the relatively small cell-size, this is not significantly (p = .095) more than the 14% one would expect from a random design.

The lacking significance for supply chain E is perhaps surprising in light of the high Spearman correlation (.51) from Table 2. However, the high level of co-production “eats up degrees of freedom” and causes the performance difference between a random and an optimal design to be very small, leaving us with less statistical power. We will see this more clearly in the supply chain level analysis presented below.

Looking beyond the individual entries at the overall pattern of results in Table 3, we see that most rows show a monotonic increase indicating that pairs are more likely to be co-produced if the adjustment frequency is higher. We could also perform a statistical test of this relationship, but since a more correct analysis takes indirect effects into account, we do not offer any analyses of direct effects beyond those reported above.

In spite of the fact that we used the aforementioned conservative test, Table 3 shows support for the hypothesis that a pair of parts is more likely to be produced by the same firm if mutual adjustments are needed on a more frequent basis. The results are also broadly based in the sense that we see the same general pattern across all the industries in spite of their apparent differences.

III.ii. Tests at the supply chain level

The analysis in Table 3 is incomplete because it only takes direct effects into account. In fact, it is closely related to the one-part-at-a-time studies criticized in the Introduction. To capture indirect effects, we need a test that takes account of the entire matrix of adjustment frequencies. To this end, we use as our measure the sum of importance-weighted internalized adjustments. This can be formally expressed as $\Sigma_{ij}z_{ij}a_{ij}$, where z_{ij} is a 0-1 indicator of co-production. (We will later see that this is the correct measure of performance if the number of firms is held constant and the variable adjustments-costs are independent of firm size.) To test the hypothesis that a disproportionately large number of adjustments are internalized, we compare the actual value of this measure against those in 100,000 random designs.[18] Following the arguments made before Table 3, we again opt to be conservative and constrain the random designs to have the same firm sizes as the actual.

The use of $\Sigma_{ij} z_{ij} a_{ij}$ assumes that we can treat the a_{ij}'s as cardinal although they in fact are ratings on a discrete seven point scale. This is a problem if the ratings are non-linear in the true adjustment frequencies. At best, it introduces noise and bias against hypotheses, and at worst it might cause the measure to rank alternative designs incorrectly. It is hard to assess the seriousness of this problem, but it is one the paper shares with most other studies based on questionnaire data.

The results of the test are presented in Table 4 below. The second column give the means of the measures ($E\Sigma_{ij} z_{ij} a_{ij}$) based on the simulations, the numbers in the fourth column ($Max_z \Sigma_{ij} z_{ij} a_{ij}$) are its theoretical maximum (the best that can possibly be done within the constraints of the actual size-distribution) and is found by integer programming.[19] The third column reports the actual measures of internalized adjustments ($\Sigma_{ij} y_{ij} a_{ij}$), where y_{ij} is the value taken by z_{ij} in the data (a 0-1 indicator of actual co-production).

[18] Another way to approach the problem is to use an entirely different statistical technique. A particularly interesting candidate may be cluster analysis, the use of which has some history in sociological and ecological studies of network effects (Frank, 1995).

[19] There are many ways to formulate the IP in question, but the essential trick is to define a set of 0/1 variables for each firm/part pair. That is, if the supply chain has a firm with 6 parts, the 36 variables in question have to sum to one.

Table 4

Sum of Internalized Adjustments by Supply Chain[1]

Car	$E\Sigma_{ij}z_{ij}a_{ij}$	$\Sigma_{ij}y_{ij}a_{ij}$	$Max_z\Sigma_{ij}z_{ij}a_{ij}$
A	161	275 *(.0001)*	395
B	142	221 *(.002)*	397
C	79	143 *(.005)*	288
D	100	195 *(.0001)*	316
E	433	468 *(.120)*	521
F	95	147 *(.017)*	329
G	68	152 *(.00000)*	269
H	59	134 *(.0001)*	251

[1] *p*-values in parentheses refer to tests relative to 100,000 randomly designed supply chains.

Consistent with the results at the pair-by-pair level, we see that the measures at the level of supply chains, except for car E, are very significant. In fact, four of the supply chains internalize more adjustments than 99,995 randomly generated designs, and G beats all 100,000. The result is also robust in the sense that all supply chains show similar patterns. So with the exception of supply chain E, there is statistical support for the claim that a disproportionate number of adjustments are managed inside firms.

One way to measure the magnitude of the effect is by comparing the performance of random designs ($E\Sigma_{ij}z_{ij}a_{ij}$), actual designs ($\Sigma_{ij}y_{ij}a_{ij}$), and optimal designs ($Max_z\Sigma_{ij}z_{ij}a_{ij}$). We can interpret ($\Sigma_{ij}y_{ij}a_{ij}$ - $E\Sigma_{ij}z_ia_{ij}$)/(Max$_z\Sigma_{ij}z_{ij}a_{ij}$ - $E\Sigma_{ij}z_{ij}a_{ij}$) as the actual "excess internalization" divided by the highest possible "excess internalization" (what would be observed if the theory explained everything and our measures were perfect). The average of this ratio is around *.33,* suggesting that one third of the forces captured by our measures are reflected in the actual design. However, the sheer size of the optimization problem raises an almost philosophical question about the use of full optimality as a benchmark. Clearly, no precise solution has been feasible until very recently and it is difficult to think that the industry, based on just 100 years of competition or experience, would have found the most efficient structure among so many possibilities. Even the very high performing supply chain for car G scores less than one half on the $(\Sigma_{ij} y_{ij} a_{ij} - E\Sigma_{ij} z_i a_{ij})/(\mathrm{Max}_z \Sigma_{ij} z_{ij} a_{ij} - E\Sigma_{ij} z_{ij} a_{ij})$ measure. While this certainly is a case where bounded rationality is a reasonable assumption, it is obviously very hard to argue for any specific benchmark other than full optimality.

III.iii. An optimization model

The results in Table 4 tells us that actual supply chain designs internalize many more adjustments than one would expect if designs were random. They do not tell us why this is. To this end, we will use simulated GMM to estimate an optimization model of supply chain design, aiming to show that it can be portrayed as the solution to a program like *(P)*. This will allow us to measure the extent to which ours is the right model of supply chain design. Specifically, we formulate a program in which adjustment-costs plus noise are minimized, and show that the former can not be ignored. One way to do this is by writing the objective function in *(P)* as $(N-n)w + \frac{1}{2}\Sigma_{i\in\mathbf{N}}[f(S_I)\Sigma_{j\in I}(\beta a_{ij} + e_{ij}) + m\Sigma_{j\notin I}(\beta a_{ij} + e_{ij})]$, where once again N is the number of parts, n is the number of firms, $w(S_b-1)$ are fixed adjustment-costs in a firm of size S_b, $f(S_b)$ are the variable adjustment-costs in a firm of the same size, m are the variable adjustment-costs in the market, a_{ij} is the frequency of mutual adjustment, and e_{ij} is normally distributed noise. The idea is now to evaluate the importance of adjustment frequencies by looking at the magnitude and significance of β. Since the theory allows the $f(S_b)$'s to be constant, we simplify a bit by restricting them to be independent of firm size. This allows us to write the objective function of *(P)* in terms of the two parameters f/w and m/w, thus reducing the dimensionality of the estimation problem. On the other hand, since this is an unnecessary restriction, we achieve the formal simplicity at the cost of estimating a coarser and presumably less well fitting model.

There are, however, significant computational barriers to even this plan. To estimate the model with simulated GMM, we start with a provisional parameter value β' (say 1) and solve P for a number of randomly drawn $\{e_{ij}\}$ matrices. After using the appropriate moment conditions to evaluate β', we repeat the procedure for a new parameter value, continuing until we find the best estimate. We might thus end up solving *(P)* more than 2000 times. But for $N = 36$, the number of feasible solutions to *(P)* is comparable to the number of seconds passed since the big bang, resulting in what presently are insurmountable computational demands. Rather than reducing the number of parts, we have once again chosen to constrain the simulated solutions to the same size-distribution of firms as that in the data, denoted by $(S_1^o, S_2^o, ..S_n^o)$. That is, we test that parts are allocated to minimize adjustment-costs under the assumption that the supply chain has to follow an exogenously imposed size-distribution. In our first two tests, reported in Tables 3 and 4, we made the same assumption in order to be conservative. It is hard to evaluate whether it is conservative in this third test, but we here have to make it for computational reasons.[20]

[20] Apart from the theoretical advantages of estimating the model without this constraint, it would also allow us to calibrate the model by estimating the relative magnitudes of fixed- versus variable

A fixed size-distribution implies that the number of firms is constant, and the objective function in *(P)* reduces to $\frac{1}{2}\Sigma_{i\in\mathbf{N}}[f\Sigma_{j\in I}(\beta a_{ij} + e_{ij}) + m\Sigma_{j\notin I}(\beta a_{ij} + e_{ij})]$. This can be expressed as $\frac{1}{2}(f-m)\Sigma_{i\in\mathbf{N}}\Sigma_{j\in I}(\beta a_{ij} + e_{ij}) + \frac{1}{2}m\Sigma_{i\in\mathbf{N}}\Sigma_{j\in\mathrm{N}}(\beta a_{ij} + e_{ij})$, where the last term is a constant and $(f-m) < 0$. So total adjustment-costs are minimized by having as many intra-firm adjustments as possible, while respecting the actual size-distribution. Consequently, *(P)* is equivalent to the following much simpler Partitioning Problem

$$\mathrm{Max}_{\mathbf{S}}\ \Sigma_{i\in\mathbf{N}}\Sigma_{j\in I}(\beta a_{ij} + e_{ij}),\ \text{s. t. } S_b = S_b^o,\ \text{for all } b. \qquad (P')$$

We can interpret *(P')* as a problem in which 36 parts have to be put into n firms of predetermined sizes in such a way that the sum of the intra-firm benefits, $(\beta a_{ij} + e_{ij})$, is maximized. If $\beta = 1$ and the variances of the e_{ij}'s were 0, the objective in *(P')* is the same measure used in the simulations underlying Table 4. Because we are holding the size-distribution constant, the β's can not give us insights into the relative magnitudes of fixed- versus variable adjustment costs. They are just weights relative to the random effects, with the values of 0 and 1 being the most natural benchmarks.

While it is much, much smaller than *(P)*, *(P')* is still a very large integer programming problem. In the typical supply chain, it has more than 10^{22} possible solutions, but can be formulated as a linear integer program with about 2000

---

adjustment costs. There are three possible ways to reduce the computational burden. The first and most direct solution is to study an industry with fewer parts. A second possibility is to use near-optimal solutions to *P*. We could conceivably use solutions that are within, say 1% of optimal, and still get close to the true parameter values. However, our experiments with this have not resulted in significant reductions in computation time. The third avenue is to define approximation by processor time and work with an inflated version of the best solution found after a fixed amount of time. We evaluated this for about 350 runs on industries B and C, and found that it worked fairly well. However, it is hard to put bounds on the degree of approximation involved.

variables and 4000 constraints. If we use a number of tricks to speed up the program, the CPLEX IP code allows us to find an optimal solution in a reasonable amount of time. Most of our individual runs take ten to fifteen minutes, although the time varies from a few seconds (for supply chain E) to several days or weeks (for supply chains F and B). The results reported below are based on a total of roughly 25,000 optimizations or one and a half years of processor time.

We estimate β for individual supply chains from the moment condition:

$$E\Sigma_{ij}[y_{ij} - Prob(x_{ij}=1|\beta^*)]a_{ij} = 0, \tag{1}$$

where y_{ij} and x_{ij} are 0-1 indicators of co-production, referring to the data and the simulations, respectively.[21] In practice, we find a measure of $Prob(x_{ij}=1|\beta')$ as the average value of $x_{ij}(\beta')$ over 100 simulations based on β'. We arrive at this average as follows: We first draw 100 independent $\{e_{ij}\}$ matrices each consisting of 630 independent draws from $N(0, \sigma^2)$, where σ^2 is the variance of the a_{ij}'s. Given a provisional β' and the first matrix, we then solve P by finding the allocation that maximizes $\Sigma_{ij}x_{ij}(\beta')[\beta' a_{ij} + e_{ij}]$ subject to the constraint that the size-distribution is identical to the actual. After solving (P') for the other 99 $\{e_{ij}\}$ matrices, we assign β' a score of $\Sigma_{100}\Sigma_{ij}\, x_{ij}(\beta')a_{ij}/100$.[22] We repeat the process for several other provisional β's, searching for a score equal to $\Sigma_{ij}y_{ij}a_{ij}$.

To estimate the model on pooled data, we sum the left hand side of (1) over the eight supply chains to arrive at the moment condition:

---

[21] For a specific $\{e_{ij}\}$, the solution to P' does not change for very small changes in β, and $\Sigma_{ij}x_{ij}(\beta^*)a_{ij}$ is a decreasing step-function. However, since we are using the expectation of this and the $\{e_{ij}\}$ are normally distributed, the left side of (1) is a decreasing and continuous function of β.

[22] Since individual optimizations of supply chain B took up to two weeks, the analysis of that is less rigorous. We used just 30 simulations for each value of β, and in some of them went with approximate solutions.

$$E\ \Sigma_{supplychains}\{\Sigma_{ij}[y_{ij} - \Sigma_{100} x_{ij}(\beta^{*})/100]\,a_{ij}\}=0. \qquad (2)$$

We can sketch an argument to the effect that the resulting estimate β^* is consistent by taking limits as the number of supply chains (R) and the number of simulations per chain (T) go to infinity. To this end, we denote the true parameter value by β^0, a sequence of estimates by $\beta_1, \beta_2,.. \beta_n,..,$ the data-matrices by $\boldsymbol{y}(1), \boldsymbol{y}(2),...\boldsymbol{y}(T),$ and the simulated solution matrices by $\boldsymbol{x}(1, \beta_n), \boldsymbol{x}(2, \beta_n), ...\boldsymbol{x}(R, \beta_n)$. If the model is correct, the data are generated by $\boldsymbol{y}(\boldsymbol{e}, \beta^0, \boldsymbol{a})$ = Argmax$\sum_{ij} (\beta^0 a_{ij}+ e_{ij})y_{ij}$, and as R grows without bound, we will have a distribution of $\boldsymbol{y}$'s that reflect β^0 (as well as σ^2 and $\boldsymbol{a}$). The simulated solutions are generated by $\boldsymbol{x}(\boldsymbol{e}, \beta_n, \boldsymbol{a})$ = Argmax$\sum_{ij} (\beta_n a_{ij}+ e_{ij})x_{ij}$, and as T grows without bound, we will have a distribution of $\boldsymbol{x}$'s that reflect β_n (as well as σ^2 and $\boldsymbol{a}$). The distribution of the $\boldsymbol{x}$'s will converge to that of the $\boldsymbol{y}$'s as β_n converges to β^0. [23]

We calculated the standard errors of the estimates by parametric bootstrapping from the asymptotic distributions.[24] Specifically, we drew a random $\{e_{ij}\}$ matrix and found a hypothetical design by solving Max$_{\mathbf{S}}\ \Sigma_{i\in\mathbf{N}}\Sigma_{j\in I}(\beta^* a_{ij} + e_{ij})$, s. t. $S_b = S_b^o$, for all b. We then treated the solution as if it was the actual and estimated a value of β for it. Doing this 100 times gave us 100 point estimates and

[23] Since there is a fixed number of zeroes and ones in each $\boldsymbol{x}$, we may also be able to show that the simulation noise average out, such that we can work with a finite T. It may also be possible, but much harder, to get results for very large N. These results would then also apply to the estimates for individual supply chains. However, this remains a question for further research.

[24] The estimation of standard errors is complicated by the fact that the left side of *(1)* is a step-function of β for any finite number of simulations, such that the estimated value of β^* is a member of an interval. Specifically, it is not clear how to find standard errors if $\Sigma_{100}\Sigma_{ij}\, x_{ij}(\beta')a_{ij}/100$ is flat over very large intervals. Since we are looking at roughly 10^{22} solutions and $\beta \in [0, 1]$, such problems would seem "unlikely", but they can not be ruled out a priori. However, we are pleased to report that the iteration processes involved in the estimations were very well-behaved. We searched over grids of length down to .01 for the industry models, and down to .001 for the pooled model. In all cases the left hand sides of *(1)* and *(2)* are strictly decreasing in β and in most cases convex. We believe that it may be possible to develop some more rigorous arguments about this, but leave them as questions for future research.

we use the standard error from that distribution as the standard error of our estimates of β. The results are reported in Table 5, where the row labeled POOLED refers to estimates with pooled data.

Table 5

Models of Max $\Sigma_{ij}x_{ij}(\beta)[\beta a_{ij} + e_{ij}]$.[1]

Car	β^*	s.e. β^*
POOLED	.314	.095
A	.40	.15
B	.29	.11
C	.27	.10
D	.37	.14
E	.26	.26
F	.21	.11
G	.35	.12
H	.33	.13

[1]Standard errors are bootstrapped.
N=5040 for the pooled model, 630 for supply chains A-H.

In spite of the fact that we estimated a one-parameter model with coarse data, the betas are, with the exception those for supply chain E, significant. In light of the very different supply chain designs documented in Table 2, the betas are also surprisingly similar. While each of the eight supply chains has solved the design problem in its own way, it appears that they all weigh the adjustment-costs to a similar degree.

Because the a_{ij}'s and the e_{ij}'s have the same variance, we can use the magnitudes of the betas to get another perspective on the influence of adjustment frequencies. We have portrayed the supply chain as maximizing sums of $\beta a_{ij} + e_{ij}$, so $\beta^* = .33$ suggests that our theory and measures capture about .33/[.33 + 1], or one fourth, of the forces going into the determination of supply chain design. The extent to which the model fits the data is surprising in light of the complexity of the optimization problem postulated by the theory. As discussed in connection with Table 4, one can reasonably question the practical possibility of full optimality and thus of its use as a benchmark. As a way to judge the possible implications of this, we looked at the implications of smaller or larger values of β. Using the actual $\Sigma_{ij}y_{ij}a_{ij}$ score as a benchmark, and focusing on the pooled estimates, the score for $\beta = 0$ (no rationality?) is .65, the score for $\beta = ½$ is 1.23, that for $\beta = 1$ is 1.51, and that for very large β (full rationality?) is 1.59.[25] So while the firms do surprisingly well, there is room for improvement.

Summarizing the results, we have provided robust and multifaceted evidence in support of the claim that the adjustment-cost theory provides a good explanation for several apparently very different supply chain designs.

IV. Comparing Supply Chain- and Firm-Level Estimation

Since the supply chain-level model is theoretically superior to a firm-level (make-or-buy) model, it is interesting to compare the empirical results from the two models. Specifically, how much more precisely could we estimate the model by using the entire matrix of adjustment frequencies, as opposed to a single row? To look at this, we represent the firm by body-in-white - a part always made by the brand name manufacturer. Since the pair-wise ratings between body-in-white and the other parts have larger variance than those in any other row, this choice

[25] The corresponding numbers for the interview data are .68, 1.39, 1.66, and 1.73.

gives the firm-level model as much information as possible and thus leads to the most conservative evaluation.[26]

To minimize confounds, we compare the supply chain-level model to an analogously constrained and estimated firm-level GMM. Specifically, if body-in-white is co-produced with s_1 other parts, (e_{1j}) is a vector consisting of 35 independent draws from $N(0, \sigma^2)$, and σ^2 is the variance of all the a_{ij}'s, we portray the firm as solving

$$\text{Max } \Sigma_j x_{1j}(\beta a_{1j} + e_{1j}), \text{ s. t. } \Sigma_j x_{1j} = s_1. \qquad (P'')$$

In close analogy to *(1),* we estimate β from the moment condition

$$E\ \Sigma_j[y_{1j} - \Sigma_{100} x_{1j}(\beta^*)/100]a_{1j} = 0. \qquad (3)$$

For each value of a_{ij}, this model gives us an estimate of $Prob\ (x_{1j}=1|a_{ij})$. To evaluate the extent to which the firm-level model fits the supply chain data, we insert these probabilities into the entire matrix, subtract $\{y_{ij}\}$, square each cell, and sum the squares to find the sum of squared residuals (SSR) over the 630 cells.

As there is little reason to estimate a firm-level model with GMM, we also estimate an (unconstrained) logit model of co-production with body-in-white. Since this model is based on a different functional form, it is a bit harder to compare with our supply chain model, but we can measure its relative performance by finding the SSR over the 630 cells as above. The SSRs from the supply chain model, the firm-level GMM model and the firm-level logit model are given in Table 6 below.

[26] If we had chosen to represent the firm by a part that is produced in isolation, the logit model would have been inestimable, while our model would be unchanged.

Table 6

Sum of Squared Supply Chain Residuals (SSR) by Alternative Models.[1]

Car	Supply Chain GMM	Firm GMM	Firm Logit
A	117.2	136.2	133.5
B	117.4	151.1	150.5
C	74.4	89.3	88.6
D	84.8	99.3	97.4
E	105.9	115.1	114.6
F	88.6	100.6	99.2
G	60.5	72.9	71.9
H	56.7	70.8	70.8

[1] All entries are based on 630 pairs of parts.

The table shows that the supply chain-level model consistently fits better than either of the two firm-level models. Since it makes more intensive use of the data, it is not surprising that the supply chain-level model does better than the firm-level GMM model. Because its theoretical advantage is larger when fewer parts are co-produced, we would expect the superiority of the supply chain-level model to differ between supply chains. The Table bears this out, as the ratio between the SSRs generally is larger in supply chains like H, G, C, F, and D where fewer parts are co-produced, and smaller in supply chain E.

As mentioned above, it is harder to compare the supply chain GMM with the logit model because the difference in data sets is confounded with a difference in functional forms. To (imperfectly) decompose the effects, we can start by comparing the two firm-level models. From Table 6, we see that the logit model fits the questionnaire data bit better, but Table 6A in the Appendix shows that the firm-level GMM fits the interview data better. As one would expect, the relative advantages of the two functional forms depend on the data sets. On the other hand, the logit model clearly does less well than the supply chain-level GMM, suggesting that any advantages tied to functional forms are overwhelmed by those associated with more intensive use of the data.

One could argue that the above model comparison is biased against the firm-level models in favor of the supply chain-level model, because only the latter is estimated on the data used for comparison. However, we will claim that this standard of comparison is the only correct one, since it is consistent with the belief that all co-production decisions follow the same logic. We can nevertheless get another take on the model comparison by evaluating model performance relative to co-production with body-in-white only. These results are given in Table 7 below.

Table 7

Sum of Squared Firm Residuals (SSR) by Alternative Models.[1]

Car	Supply Chain GMM	Firm GMM	Firm Logit
A	7.41	8.52	8.29
B	7.31	5.01	4.75
C	7.25	6.59	6.35
D	7.26	8.01	7.88
E	3.57	3.62	3.53
F	8.58	8.45	7.95
G	6.13	7.29	7.14
H	6.36	7.15	7.06

[1] All entries are based on 35 pairs of parts.

Since the two firm-level models are estimated on that data only, this measure does not penalize them for over-fitting and will be more favorable to them. Even so, the supply chain-level model still outperforms the two firm-level models in more than half of the industries. We conclude that supply chain-level estimation offers significant advantages over firm-level (make-or-buy) estimation, but admit that it uses more data and poses non-trivial computational difficulties.

V. Discussion

Using original data that map directly to the theory, we have presented several successively deeper tests of the proposition that the firm is a low variable-, but high fixed cost way to govern adjustments. Our results are strong and robust across alternate specifications and supply chains. By taking a supply chain-level perspective, we have been able to avoid the mis-specifications inherent in make-or-buy estimates and extract more information from the data, especially in the cases where many parts are out-sourced. Consistent with this, we found that our optimization model fits better than a firm-level model, both theoretically and in this dataset. Our innovative testing strategy is not tied to the adjustment-cost theory and it should be possible to apply it to other theories of the firm as well.

We freely admit that the empirical analysis can and should be improved. Our data are very good and we particularly encouraged by the .915 correlation between our two measures of adjustment frequency. However, we cannot generally rule out that "the frequency with which there needs to be a mutual adjustment", is picking up something other than adjustment frequency.[27] As other innovative testing strategies, ours opens up a new territory with no prior work. It is hard to say the first and last word on an important topic. However, based on the

[27] As a preliminary check in this direction, we asked our expert for a rough matrix with ratings of the "magnitudes" of mutual adjustments, and can report that analyses with that produced much weaker results than those with frequencies.

importance of the theoretical questions, as well as the strength and robustness of our results, we believe that the paper is a worthwhile and useful starting point.

An important area for future research is to test competing theories and combinations of theories. In the simplest possible execution, each theory could be represented by a matrix measuring the predicted degree of contracting difficulty between each pair of mega-parts. For example, a test of the property rights theory would rely on a matrix of co-specialized assets. We could then run a horse race between the different theories based on the fit of the corresponding optimization models. While the reader will appreciate the computational challenges of estimating multiple parameters on the present data-set, our techniques transfer readily to simpler industries. It will thus be very desirable to develop alternative data-sets.

Another interesting avenue for further research is to deepen the analysis of the adjustment-cost theory. We have measured adjustments in terms of frequency only – treating them all as the same. However, one would expect that different types of adjustments carry different expected costs. In the extreme case where a small set of identical adjustments may be repeated, the theory predicts that the parties will agree on ex ante pricelists – lowering variable costs of adjustment in the market (Wernerfelt, 1997). It would be interesting, but challenging, to develop a measure of adjustment difficulty and use it to moderate the frequencies.

It is possible to interpret our results as reflective of endogeneity with the idea being that there are more or fewer mutual adjustments between two parts *because* they are or are not co-produced. However, we are not too worried about this. First, since we collected this measure in two different ways, both would have to be subject to the problem in essentially the same way. Secondly, all data collected explicitly differentiated the needs for mutual adjustments from actual adjustments. We took pains to ask the experts about "the frequency of *needs* for mutual adjustments". Thirdly, our theory predicts that adjustment between co-produced parts is cheaper. In the interest of simplicity, we derived *P* under the assumption that all needed adjustments are made. But without this assumption, we would find that more of the needed adjustments are implemented within than between firms. It is certainly possible to question our point estimates based on this line of reasoning, but the argument relies on the premise is that the theory is correct. So to the extent that reverse causality is driven by differences in adjustment-costs, we are not too unhappy about it.

In closing, the robust patterns in the data presented here should give pause to those who a priori reject the importance of ex post adjustment in the theory of the firm. We recognize that much work remains to be done, and hope that the approach taken here will prove fruitful in further testing of this and other theories of the firm, as well as in theories of more general network structures.

References

Baldwin, Carliss Y., and Kim B. Clark, Design Rules: The Power of Modularity, Cambridge MA.: The MIT Press, 2000.

Bradley, Peter, "Juggling Tasks: It is just Another Buying Day", Purchasing, Available from the second author on request.

Busse, Meghan, Jorge Silva-Risso, and Florian Zettelmeyer, "$1000 Cash Back: The Pass-Through of Auto Manufacturer Promotions", American Economic Review, 96, no. 4, September, pp. 1253-70, 2006.

Coase, Ronald H., "The Nature of the Firm", Economica New Series, 4, no. 16, pp. 386-405, 1937.

Eppinger, Steven D., "Model-Based Approaches to Managing Concurrent Engineering," Journal of Engineering Design, 2, no. 4, pp. 283-90, 1991.

Frank, Kenneth A., "Identifying Cohesive Subgroups", Social Networks, 17, no. 1, pp. 27-56, 1995.

Grossman, Sanford, and Oliver D. Hart, "The Costs and Benefits of Ownership," Journal of Political Economy, 94, no. 4, August, pp. 691-719, 1986.

Hart, Oliver D., "Reference Points and the Theory of the Firm, Economica, 75, no. 299, pp. 404-11, 2008.

Hart, Oliver D., and John Moore, "Contracts as Reference Points", Quarterly Journal of Economics, 123, no. 1, February, pp. 1-48, 2008.

Holmstrom, Bengt, and Paul Milgrom, "The Firm as an Incentive System," American Economic Review, 84, no. 5, September, pp. 972-91, 1994.

Lafontaine, Francine, and Margaret Slade, "Vertical Integration and Firm Boundaries: The Evidence", Journal of Economic Literature, 45, no. 3, pp. 629-85, 2007.

Monteverde, Kirk, and David Teece, "Supplier Switching Costs and Vertical Integration in the U.S. Automobile Industry", Bell Journal of Economics, 13, no. 1, Spring, pp. 206-13, 1982.

Nagaoka, Sadao, Akira Takeishi, and Yoshihisa Noro, "Determinants of Firm Boundaries: Empirical Analysis of the Japanese Auto Industry from 1984 to 2002", Journal of the Japanese and International Economies, 22, no. 2, pp. 187-206, 2008.

Novak, Sharon, and Steven D. Eppinger, "Sourcing by Design: Product Complexity and the Supply Chain," Management Science, 47, no. 1, January, pp. 189-204, 2001.

Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model", Econometrica, 50, no. 1, January, pp. 97-109, 1982.

Simester, Duncan I., and Marc Knez, "Direct and Indirect Bargaining Costs and the Scope of the Firm," Journal of Business, 75, no. 2, April, pp. 283-304, 2002.

Simon, Herbert A., "A Formal Theory of the Employment Relationship", Econometrica, 19, no. 3, July, pp. 293-305, 1951.

Steward, D. V., "The Design Structure System: A Method for Managing the Design of Complex Systems," IEEE Transactions in Engineering Management, 28, no. 3, pp. 71-84, 1981.

Watson, Joel, "Alternating-Offer Bargaining with Two-Sided Incomplete Information", Review of Economic Studies, 65, July, pp. 573-94, 1998.

Wernerfelt, Birger, "On the Nature and Scope of the Firm: An Adjustment-Cost Theory," Journal of Business, 70, no. 4, October, pp. 489-514, 1997.

Williamson, Oliver E., "Transaction-cost Economics: The Governance of Contractual Relations," Journal of Law and Economics, 22, no. 4, October, pp. 223-61, 1979.

Appendix: Results Based on Interview Data

Table A-1

Adjustment Frequencies from Interviews and Co-production Data for Supply Chain A[28]

	6	3	2	3	2	1	1	3	2	0	3	3	6	0	3	4	3	5	6	6	0	0	2	0	0	5	0	0	4	3	2	2	3	3	0
0		5	2	0	0	6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
0	0		0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3
0	0	1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0		6	0	0	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1		0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0		0	2	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0		4	4	2	1	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	0	0		6	3	2	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	3	0	0	0
0	0	0	0	1	1	1	0	0	1		3	0	0	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	0		4	6	6	5	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0		5	1	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0		5	5	0	0	0	0	0	0	0	0	2	0	4	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0		4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0		0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1		5	2	6	6	4	2	0	2	2	0	2	3	3	2	3	2	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1		6	5	5	3	3	2	4	4	0	6	6	6	0	5	2	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1		3	3	0	1	4	0	1	0	3	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1		6	5	3	1	2	2	0	4	5	2	0	4	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1		5	3	1	2	2	0	4	5	2	0	4	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1		0	0	3	3	0	3	3	0	0	3	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		2	1	1	0	0	0	1	0	1	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0		0	0	0	0	0	2	0	0	0	0	0	1
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1		5	1	3	5	1	0	4	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	5	6	4	0	4	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0		0	1	0	0	3	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1		6	2	2	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		3	4	4	4	2	2	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0		1	0	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		5	1	6	6	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	0		0	3	3	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	1	0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	0	1	0		6	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		0
1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	0	1	0	1	0	

[28] The highlighted 3 in the upper left corner refers to the adjustment frequency between "body-in-white" and "headlining", while the highlighted 0 tells us that the two are produced by different firms.

Table A-2

Descriptive Statistics by Supply Chain[1]

Car	Fraction co-produced	Number of parts co-produced	*r*(co-production, adjustment frequency)
A	.29	19, 5, 2, 2, 2	.52
B	.26	16, 9, 4, 2	.49
C	.14	13, 5, 2, 2, 2	.57
D	.18	15, 3, 3, 2, 2, 2, 2	.53
E	.79	32	.48
F	.17	14, 5, 3, 3, 2, 2	.53
G	.12	12, 4, 3, 2, 2, 2	.61
H	.11	11, 4, 3, 3, 2	.60

[1]N=630.

Table A-3

Probability of Pairwise Co-production by Adjustment Frequency[1]

Car	0	1	2	3	4	5	6
A	.235 *(.002)*	.385 *(.218)*	.386 *(.134)*	.410 *(.114)*	.478 *(.045)*	.476 *(.061)*	.619 *(.005)*
B	.226 *(.016)*	.231 *(.697)*	.182 *(.917)*	.385 *(.073)*	.217 *(.754)*	.571 *(.004)*	.667 *(.001)*
C	.116 *(.017)*	.192 *(.321)*	.159 (.455)	.256 *(.067)*	.174 *(.424)*	.238 *(.182)*	.333 *(.033)*
D	.154 *(.028)*	.115 *(.866)*	.182 (*.567)*	.205 *(.199)*	.217 *(.402)*	.429 *(.010)*	.476 *(.003)*
E	.768 *(.150)*	.808 *(.506)*	.795 *(.525)*	.821 *(.448)*	.870 *(.249)*	.905 *(.173)*	.905 *(.186)*
F	.156 *(.099)*	.192 *(.468)*	.045 *(.998)*	.231 *(.247)*	.130 *(.801)*	.429 *(.007)*	.476 *(.003)*
G	.094 *(.009)*	.000 *(1.00)*	.045 *(.978)*	.282 *(.008)*	.130 *(.555)*	.429 *(.001)*	.476 *(.001)*
H	.083 *(.015)*	.115 *(.542)*	.091 *(.708)*	.205 *(.068)*	.174 *(.224)*	.238 *(.070)*	.286 *(.025)*

[1] *p*-values in parentheses refer to tests relative to 100,000 randomly designed supply chains.

Table A-4

Sum of Internalized Adjustments by Supply Chains[1]

Car	$E\Sigma_{ij}z_{ij}a_{ij}$	$\Sigma_{ij}y_{ij}a_{ij}$	$Max_z\Sigma_{ij}z_{ij}a_{ij}$
A	162	264 *(.001)*	407
B	143	231 *(.001)*	420
C	80	132 *(.012)*	312
D	101	174 *(.004)*	340
E	436	476 *(.097)*	526
F	96	153 *(.011)*	356
G	69	154 *(.0000)*	295
H	60	112 *(.005)*	274

[1] *p*-values in parentheses refer to tests relative to 100,000 randomly designed supply chains.

Table A-5

Models of Max $\Sigma_{ij}x_{ij}(\beta)[\beta a_{ij} + e_{ij}]$.[1]

Car	β^*	s.e. β^*
POOLED	.276	.069
A	.33	.13
B	.31	.08
C	.23	.09
D	.26	.11
E	.28	.36
F	.23	.09
G	.32	.10
H	.25	.10

[1]Standard errors are bootstrapped.
N=5040 for the pooled model, 630 for supply chains A-H.

Table A-6

Sum of Squared Supply Chain Residuals (SSR) by Alternative Models.[1]

Car	Supply Chain GMM	Firm GMM	Firm Logit
A	115.2	131.8	132.8
B	116.3	149.8	149.3
C	73.7	92.1	93.0
D	87.9	98.2	98.5
E	114.6	107.0	107.1
F	88.2	96.1	96.6
G	62.4	68.2	68.6
H	59.2	66.7	66.1

[1] All entries are based on 630 pairs of parts.

Table A-7

Sum of Squared Firm Residuals (SSR) by Alternative Models.[1]

Car	Supply Chain GMM	Firm GMM	Firm Logit
A	7.29	8.26	8.14
B	6.77	4.48	4.54
C	7.16	5.72	5.66
D	7.78	7.42	7.31
E	3.88	3.57	3.50
F	8.69	7.71	7.73
G	6.51	6.83	6.74
H	6.98	6.92	6.80

[1] All entries are based on 35 pairs of parts.