

# Measuring Imputed Cost in the Semiconductor Equipment Supply Chain

Morris A. Cohen • Teck H. Ho • Z. Justin Ren • Christian Terwiesch

*Operations and Information Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

*Marketing Group, Haas School of Business, University of California, Berkeley, California 94720*

*Operations and Technology Management Department, School of Management, Boston University, Boston, Massachusetts 02215*

*Operations and Information Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

*cohen@wharton.upenn.edu • hoteck@haas.berkeley.edu • ren@bu.edu • terwiesch@wharton.upenn.edu*

---

We consider the order-fulfillment process of a supplier producing a customized capital good, such as production equipment, commercial aircraft, medical devices, or defense systems. As is common in these industries, prior to receiving a firm purchase order from the customer, the supplier receives a series of shared forecasts, which are called “soft orders.” Facing a stochastic internal manufacturing lead time, the supplier must decide at what time to begin the fulfillment of the order. This decision requires a trade-off between starting too early, leading to potential holding or cancellation costs, and starting too late, leading to potential delay costs. We collect detailed data of shared forecasts, actual purchase orders, production lead times, and delivery dates for a supplier-buyer dyad in the semiconductor equipment supply chain. Under the assumption that the supplier acts rationally, optimally balancing the cancellation, holding, and delay costs, we are able to estimate the corresponding imputed cost parameters based on the observed data. Our estimation results reveal that the supplier perceives the cost of cancellation to be about two times higher and the holding costs to be about three times higher than the delay cost. In other words, the supplier is very conservative when commencing the order fulfillment, which undermines the effectiveness of the overall forecast-sharing mechanism.

*(Imputed Cost; Supply Chain Management; Semiconductor Equipment)*

---

## 1. Introduction

Many firms selling expensive customized capital goods such as production equipment, commercial aircraft, medical devices, or defense systems face an order-fulfillment dilemma. On the one hand, their customers expect them to exhibit a high degree of responsiveness, requiring product delivery within an aggressive lead time. On the other hand, the customized nature of these products makes it risky for suppliers to keep subsystems or even finished goods in inventory, leading to lengthy and stochastic manufacturing lead times. To resolve this dilemma,

they routinely begin procurement or even production based on forecasted orders, as opposed to waiting for firm purchase orders from their customers. Such forecasted orders, sometimes also referred to as “soft orders,” are based on customers’ purchase intent as typically revealed to the supplier’s marketing and sales function.

There exists, however, a substantial difference between a forecasted (soft) order and a firm purchase order. Responding to changes in their technology and market environments, customers may decide to revise the shared forecasted orders, leading to either

changes in purchase quantity, delays in the requested delivery dates, or cancellations of the orders. This raises an interesting question as to how a supplier should respond to the shared forecasts in such a volatile environment. Specifically, the supplier must decide on how to deal with the preliminary nature of a soft order, and—more operationally—define a point in time (a level of information quality) at which to start the corresponding fulfillment process. The supplier can start early, facing the risk of the order not materializing (cancellation cost) or the equipment being finished too early (holding cost), or can delay the starting time until more information becomes available, thereby facing the risk of being late (delay cost). Put differently, the supplier can “wait and be late or rush and be wrong,” neither of which seems desirable.

In this article, we report on an empirical study of forecast sharing related to the acquisition of customized production equipment for the manufacturing of semiconductors. Semiconductor equipment is a prototypical example of the industry characteristics outlined above. Customers, who are responding to the turbulent environment they face in the demand for their end products, press for short customer lead times, requiring product delivery within three months or less. At the same time, the complexity and degree of customization of the equipment causes manufacturing lead times to be long and stochastic, ranging from several months to an entire year.

From an overall 143 soft orders we collected, 43 were cancelled and therefore never materialized as business for the supplier. From the remaining 100 soft orders, which ultimately were translated into firm purchase orders, another 76 experienced changes with respect to the delivery date. Taking the perspective of a supplier of customized semiconductor production equipment, we develop a formal model addressing the trade-off between an early start of the order-fulfillment process (leading to potential cancellation and holding costs) and a delayed start until more information has become available (leading to a potential delay cost).

How this trade-off is resolved depends on the perceived cost structure of the supplier. While traditionally the supply chain literature has taken

these cost parameters as exogenously given and then searched for the optimal operational decision, we take a “reverse engineering” approach. Based on empirical observation of the supply chain over time, including detailed data of shared forecasts, actual purchase orders, production lead times, and delivery dates, and on the assumption that the supplier is a rational actor, we reconstruct the cost parameters that explain the observed supply chain behavior. Our estimation results reveal that the supplier perceives the cost of cancellation to be about two times higher and the holding costs to be about three times higher than the delay cost. In other words, the supplier is very conservative when commencing the order-fulfillment process, which undermines the effectiveness of the overall forecast-sharing mechanism.

Specifically, this paper makes three contributions. First, we formally quantify the effectiveness of an order forecast-sharing system in the semiconductor equipment supply chain and show that it is less effective than one would expect. Fearing order cancellation, the supplier, by and large, ignores the preliminary forecast information. Second, to the best of our knowledge, this research is, within the field of supply chain management, the first to not only present and analyze a mathematical model, but also to empirically estimate its cost parameters. Thus, in contrast to earlier studies, the cost parameters in our research are not exogenously defined but are empirically estimated for the semiconductor equipment supply chain. This novel approach complements the growing literature of supply chain management by introducing an econometric dimension. Our results can be used by buyers who want to infer the cost parameters used by their suppliers in responding to the shared forecast information. Third, we estimate the cost parameters by adopting a “structural approach” in assuming that the supplier is rational. Consequently, the estimated parameters are robust in the face of policy changes in the presence of rational expectations (Lucas 1976).

The remainder of this article is organized as follows. Section 2 describes the semiconductor equipment industry, including the buyer-supplier relationship under investigation. Section 3 reviews the relevant literature. We present a formal optimization

model that captures the essence of the supplier's decision problem in §4, and §5 describes the data and the underlying econometric framework. Section 6 discusses the estimation results. We validate the model in §7 and conduct a policy scenario analysis in §8. Section 9 concludes the paper.

## 2. Semiconductor Equipment Supply Chain

The demand for semiconductor production equipment is triggered by the demand for the chip supplier's end products, such as microprocessors or memory chips. As illustrated by the recent downturn in these markets, the demand from PC suppliers exhibits a high degree of uncertainty, leading to wild adjustments in sales forecasts within a relatively short period of time. Market forecasts are done monthly or quarterly at large chip manufacturers for a time period of two to five years into the future and are regularly updated following a rolling-horizon principle.

These product-level demand forecasts are translated into capacity requirements and allocated to either existing or new wafer fabs. If the allocated demand is not supported by available capacity, new equipment is ordered. The ordering of the equipment requires chip manufacturers to consider the peak-load capacity requirements and changes in the technical process specifications of new chip generations.

While the translation of product demand into equipment orders seems straightforward, two factors make this computation extremely complex. First, the semiconductor industry is very capital intense, and the capital expenditures for new production equipment are the single largest item on a company's earning statement. For example, the industry leader Intel Inc. spent \$5 billion for equipment acquisitions in the year 2000 alone. Given this magnitude of capital expenditure, even minor underutilization of equipment can have a dramatic financial impact.

Second, while most semiconductor equipment in the world is operated 24 hours a day for 7 days a week, the actual availability of a specific piece of equipment can be lower. Equipment becomes unavailable as a result of machine breakdowns, required qualification procedures, engineering trials,

preventive maintenance, etc. Moreover, semiconductor manufacturing is a highly yield-driven process, associated with substantial scrap and the need for rework. While all these factors reduce the average availability, they also introduce substantial variability into capacity planning, aggravating problems resulting from the demand uncertainty.

### 2.1. Equipment-Acquisition Process

Once the capacity planning process has generated requests for additional pieces of equipment, an elaborate tool acquisition process commences. This process includes three stages, namely, forecast sharing, manufacturing, and installation.

During the forecast-sharing stage, the chip manufacturer (buyer) creates a forecasted order (soft order), which is shared with the equipment manufacturer (supplier) via an online collaboration system. This soft order includes the tool's specifications and the requested delivery date (RDD). This soft order, however, is merely preliminary information—as opposed to a final purchase order—because in the presence of market and capacity uncertainties, the buyer does not want to commit to an order at such an early stage. The buyer can, after getting more information about his market demand and production yields, decide to (a) cancel the order, (b) to move it to another date, or (c) leave the soft order unchanged.

The supplier becomes aware of the buyer's purchase intent, both through the online information system and through direct customer interaction from its sales and marketing department. At some point she needs to initiate the production of the tool, which includes procurement of subsystems from second-tier suppliers and the entry of the order into the production schedule ("slotting").

The supplier faces a difficult situation, because starting the order too early can lead to holding and cancellation costs, while starting the order too late can lead to late-shipment costs. The typical manufacturing lead time of the supplier ranges between three and five months. The lead time, however, exhibits a significant variability as a result of differences in product mix going through the supplier's facility, changes in equipment demand, process generation,

and/or uncertainty in lead times from the second-tier suppliers.

Finally, the tool is shipped to the corresponding fab, where it is installed and must then move through an elaborate qualification process before it can produce commercial output. The overall equipment-acquisition cycle is illustrated by Figure 1. In total, the equipment-acquisition cycle is approximately one year. Some tools, especially in the lithography domain, can take even longer.

Note that this equipment-acquisition process applies only to tools that have already been developed and proved their technical feasibility. The buyer uses a different contract for the supplier during the development of a new piece of production technology. One important difference between such development contracts and the procurement contracts as outlined above is that the buyer might fund more than one supplier for the development of a single tool type. In contrast, once the equipment specification is established, the buyer usually switches to a single sourcing approach.

## 2.2. Quality of Exchanged Information: The Supply Chain Dyad

The authors conducted a joint research project with a major chip manufacturer and one of the largest equipment suppliers. After documenting the overall equipment-acquisition process as outlined above, we collected detailed data on the duration of each of the three stages for a total of 100 tool orders.

Among other data, we collected the forecasted tool orders as shared by the buyer with the supplier via the online collaboration system over an extended period of time. The buyer provided a quarterly forecast of how many tools he planned to acquire in each of the coming seven quarters. These forecasts were updated following a rolling-horizon principle. For example, in quarter 4 of 1998, the buyer forecasted her demand for quarters Q1 1999 to Q3 2000. In quarter 1 of 1999, a certain number of tools were purchased and new forecasts for quarters Q2 1999 to Q4 2000 were placed.

Figure 2 depicts the forecasts as provided by the buyer in the quarters Q1 1999 to Q1 2001. Each of

Figure 1 The Equipment-Acquisition Cycle

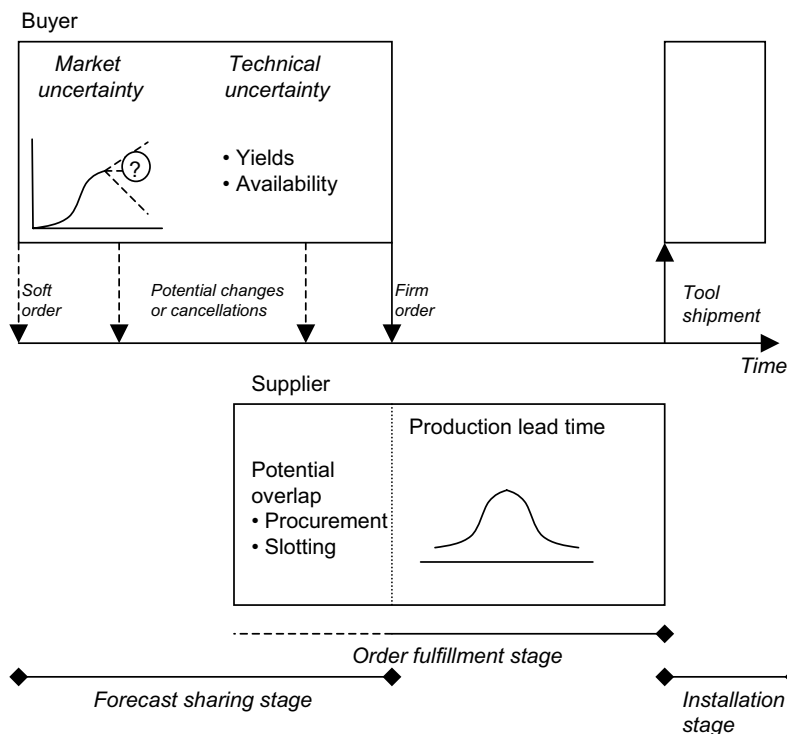
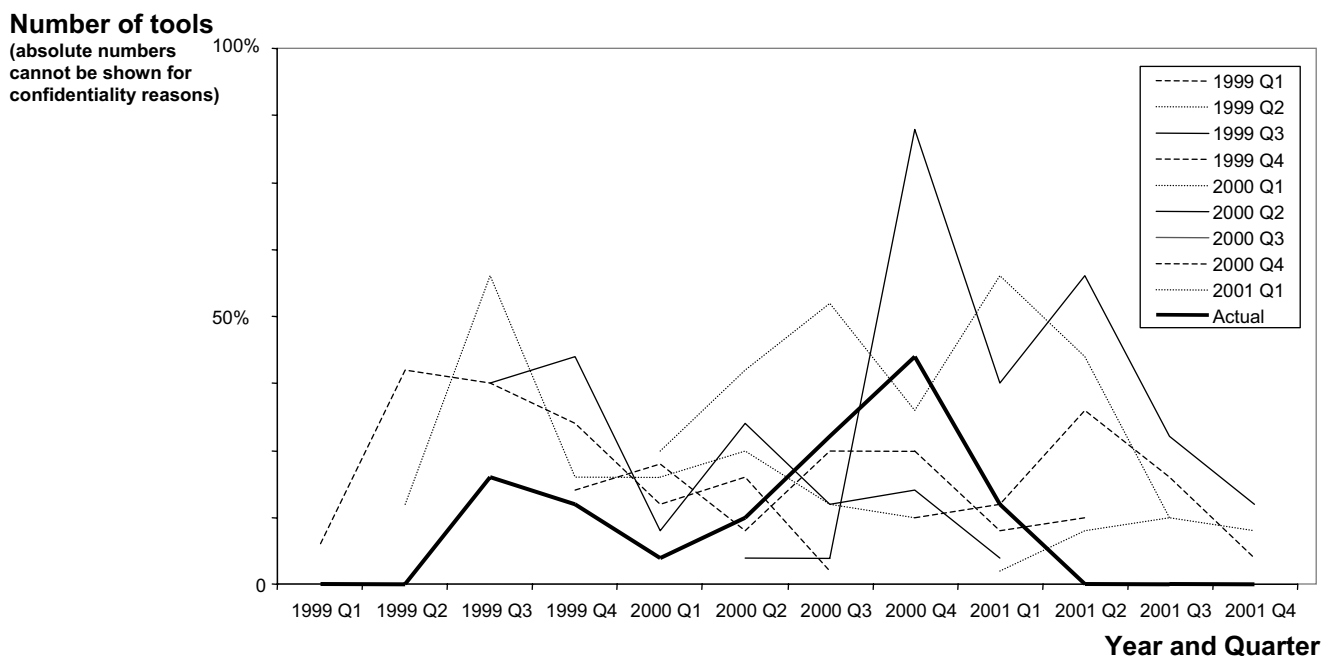


Figure 2 Forecasted (Soft) Orders vs. Actual Orders



these shared forecasts is a time series consisting of the seven quarters included in the relevant forecast window. Figure 2 also contrasts the forecasts with the actual tool purchases. We can make two interesting observations. First, the forecasts vary widely, both over time (what is forecasted in, e.g., Q1 2000 for the time period of Q2 2000 to Q4 2001) and from one forecast to the next (e.g., what is forecasted in Q2 1999 for Q4 1999 versus what is forecasted in Q3 1999 for Q4 1999). The former is a consequence of the life cycle of process generations and the associated need for capacity expansion. The latter is primarily a result of the inherent uncertainty in the industry, especially with respect to the height of the peak demand in the product life cycle.

Second, we see that—on average—the buyer forecasted for more tool purchases than he ended up committing to ultimately. In other words, there are significantly more soft orders than hard orders, leading to numerous order cancellations. In our data set, 30% of the soft orders were cancelled. This reflects the cost structure of the buyer: Forecasting too few can lead to equipment shortages and potential production losses of the entire wafer fab with a

substantial negative financial impact. Forecasting too many, however, does not directly cost the buyer. This is so because the risk of producing the equipment without having the demand for it is entirely borne by the supplier.

### 2.3. Research Goals

While a systemic inflation of forecasts does not necessarily lead to out-of-pocket costs to the buyer, it can have negative implications on the supplier's perception of the buyer's credibility. This in turn can hurt overall supply chain performance, and thereby—albeit indirectly—the buyer.

The objective of this article is to measure how the semiconductor equipment supplier we studied perceived the cost of cancellation and holding relative to the cost of late shipment. While previous research on coordination and contracting in supply chains has emphasized the importance of forecast sharing and the risks associated with losing credibility, we provide—to our knowledge—the first empirical study to demonstrate these effects econometrically based on actual supply chain behavior opposed to anecdotal evidence. Our results are of direct managerial

relevance to the semiconductor equipment supply chain because they demonstrate that the supplier does not perceive the shared forecasts as credible, which makes the process of forecast sharing, by and large, ineffective.

### 3. Related Literature

Lee et al. (1997) provide one of the earliest academic discussions of problems related to soft orders and their cancellation. They refer to such orders as “phantom orders,” defined as high forecasts of future demand that never materializes, and see them as a key contributor to the bullwhip effect in supply chains.

Problems relating to phantom orders and overly optimistic forecasts have frequently made their way into the business news. For example, Zarley and Damore (1996) discuss how PC manufacturers suspected that their customers (distributors) placed phantom orders. As a result, these manufacturers frequently produce only a fraction of the quantity specified in the demand forecast. A comparable situation occurred in the cellular phone industry in the 1994 Christmas season. Motorola experienced significant overordering by customers concerned with a potential capacity shortfall (*Business Week* 1995).

Similar problems were experienced at Boeing, which had difficulties in increasing its production of 747s due to parts shortages. Boeing’s large supplier base apparently did not trust the company’s optimistic demand forecast (indicating a strong growth in 1997) and therefore could not fill Boeing’s large orders. Only one year later, though, following the Asian financial crisis, the supplier’s conservatism proved to be a wise decision (Cole 1997a, b; Biddle 1998).

Boeing not only experienced the problem of forecast credibility with its suppliers, but—at the same time—also with their customers. For example, in May 2001, Northwest Airlines cancelled a soft order of 23 Boeing 737 jets, which led Boeing to fall behind Airbus for the first time (*Flight International* 2001).

In addition to this anecdotal evidence on problems related to forecast sharing, the recent academic literature of supply chain management includes

several articles providing game-theoretical models of this behavior.<sup>1</sup> Closest to our study, Cachon and Lariviere (2001) distinguish between forecast-sharing contracts with forced compliance and voluntary compliance. Under forced compliance, the supplier’s behavior can be observed and compared to the contracted behavior with a level of detail sufficient for any deviations to be brought up in court. In this case, all market power rests with the buyer, and he will therefore design a contract that maximizes his rent derived from the relationship.

Under voluntary compliance, the supplier’s behavior is not fully verifiable. This might be a result of some stochastic element in the environment or within the supplier’s operations. In such a situation, courts cannot enforce contract compliance because the supplier can always argue that she attempted to comply but failed because of some random event.

The situation analyzed in this article is one of voluntary compliance. While there are detailed contracts written between buyer and supplier, they are very hard to enforce. For example, if the supplier is not able to meet the requested delivery date, she could easily find reasons outside her control to explain this. Examples include parts shortages from the second-tier supplier, changing delivery dates from the buyer, and/or other external events.

A second stream of research that is relevant to our work relates to the sharing of preliminary information, which is a common practice in product development teams, especially in those proceeding concurrently. Similar to the supplier in our study, who initiates the order-fulfillment process prior to receiving a firm purchase order from the customer, development teams frequently begin their work on a new product prior to receiving detailed design specifications from the customer and/or from the market research department.

In this line of research, Krishnan et al. (1997) and Loch and Terwiesch (1998) model the situation faced by a concurrent engineering team, where an information-receiving development activity must

<sup>1</sup>Chen (2004) provides an excellent review of research in the area of sharing information within a supply chain. Our paper falls into its category of sharing downstream demand forecast information.

decide on how to rely on the preliminary information provided by the information sender. The information receiver always wants to start early, in an attempt to gain from parallel task execution. Starting early, however, uses a lower quality of information and thus has a higher likelihood of costly rework. Consequently, the information receiver faces the dilemma “rush and be wrong or wait and be late” (Terwiesch et al. 2002). This fundamental trade-off is similar to the supplier’s problem as described above and formalized below. A similar trade-off is also found in the development of new products (see for example, Cohen et al. 1996).

#### 4. Model Formulation

Our objective is to estimate the unobservable costs of cancellation, holding, and delay used by the supplier. We had observed the supply chain dyad, including shared forecasts and final shipments, over an extended period of time. In addition, we had also obtained data about the supplier’s internal manufacturing lead times. Under the assumption that the supplier would behave rationally in balancing the three cost elements, we wanted to impute values for the costs that explained observed supply chain behavior.

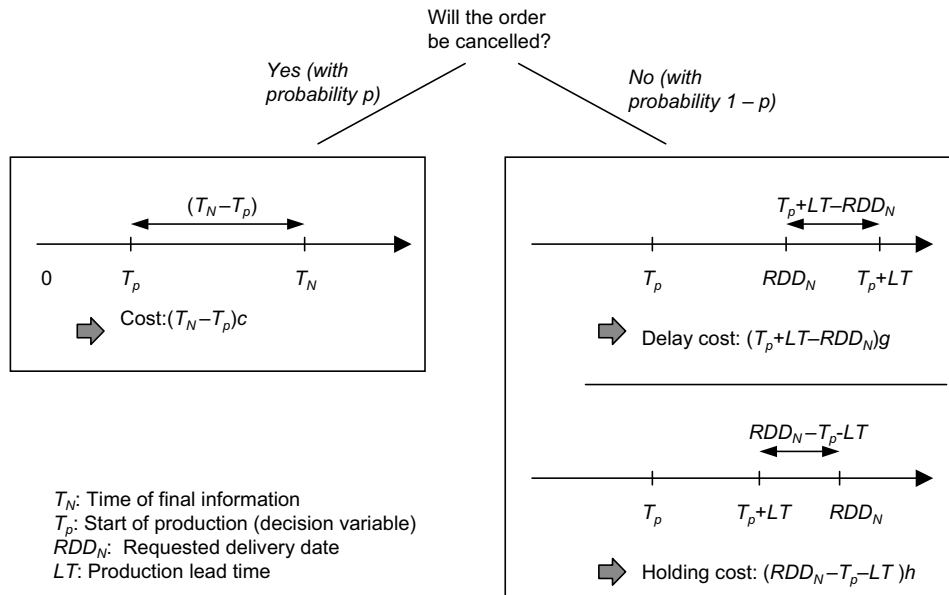
The unit of analysis in both our model and the empirical analysis is an order for a single piece of production equipment. The goal of the supplier is to choose a time to commence work on a given order so as to minimize the total costs.

##### 4.1. The Supplier’s Problem

Consider a time line (see Figure 3) starting at the point in time when the first soft order is received by the supplier ( $t = 0$ ). Associated with this first, preliminary order is a requested delivery date ( $RDD$ ), which is potentially refined by the buyer over time. At some point in time the uncertainty inherent in the soft order is resolved. Define this point in time as  $t = T_N$ . At  $T_N$ , the tool delivery is requested with a firm delivery date for  $t = RDD_N$ , or the order is cancelled, in which case we define  $RDD_N$  as infinity.

The supplier faces the following problem when deciding about the time  $T_p$  at which she begins the fulfillment process on a—potentially soft—order. Specifically, she faces two types of uncertainty: market uncertainty and uncertainty in manufacturing lead time. Market uncertainty includes the probability that the order is cancelled, which we will label as  $p$ , as well as any potential changes in the requested delivery date  $RDD_N$ .

Figure 3 Variable Definitions and Basic Cost Drivers



From the operations side, the supplier faces uncertainty in the manufacturing lead time. This may result from traditional lead-time variability in their job-shop, such as production environment, changes in product mix or production volume, and/or from variability in delivery lead times for subsystems that are ordered from second-tier suppliers.

The supplier must trade off the cost of beginning too early (cancellation and holding) with the risk of producing too late (delay cost). The problem resembles a traditional newsvendor problem, which occurs in the time domain as opposed to the usual quantity domain. Define cancellation cost,  $c$ , as the cost incurred by the supplier per unit of time that an order spends in production and later is cancelled. Cancellation costs include the cost for labor and material for an order that the supplier has started to work on. This includes component orders to second-tier suppliers that the supplier has to either return (at a charge) or hold in inventory. Cancellation costs also include the opportunity cost of not using the capacity for other orders.

Define holding cost,  $h$ , as the cost incurred by the supplier per unit of time that the tool is produced prior to the date the customer actually needs it,  $RDD_N$ . Holding cost is driven primarily by the capital cost for the expensive equipment, but also includes the cost of storing the equipment. Finally, define delay cost,  $g$ , as the cost incurred by the supplier per unit of time that the actual delivery date exceeds the  $RDD_N$ , i.e., for the shipment being late. Note that because our model takes the perspective of the supplier, these delay costs do not capture the delay cost of the buyer, who potentially loses the output of an entire production line as a result of a late shipment. Although the buyer does not face an immediate threat of substitution in the case of late deliveries (due to the single-sourcing policy adopted by the buyer), poor delivery performance will impact the supplier's likelihood of receiving contracts for future tool types and technologies.

The various cost components, including cancellation cost, holding cost, and delay cost, are illustrated in Figure 3. While it might be possible to measure holding cost from the cost of capital, machine price, and depreciation, the other two costs (delay cost and

cancellation costs) are intangible and hard to measure directly.

With these cost parameters, we can state the supplier's decision problem as follows:

$$\begin{aligned} \text{Min}_{T_p} E(\text{Total cost}) \\ = pc(T_N - T_p)^+ + (1-p)\{h[(RDD_N - LT) - T_p]^+ \\ + g[T_p - (RDD_N - LT)]^+\}, \quad (1) \end{aligned}$$

where  $(\cdot)^+$  denotes  $\text{Max}(\cdot, 0)$ . In the equation above, the first term denotes the expected cancellation cost, while the second and the third terms capture expected holding cost and expected delay cost, respectively. We assume that all three costs are linearly increasing with time. This assumption is necessary to obtain closed-form solutions that make the estimation easier. The assumption could be relaxed if the objective were to provide operational support for the supplier concerning the decision of when to begin work on a forecasted order. In such a context, one would have to collect additional data about the detailed shape of the cost accumulation functions. While holding costs are indeed likely to accumulate linearly in practice, the shape of the cancellation cost curve is likely to be lumpy. For example, upon starting on a forecasted order, the supplier is likely to experience an overproportionally high upfront cost for procurement of components.

Define  $S = RDD_N - LT$ , and let the cumulative distribution for the new random variable be  $F(S)$ . Moreover, let the distribution for  $T_N$  be  $G(\cdot)$ . Equation (1) can be rewritten as

$$\begin{aligned} \text{Min}_{T_p} E(\text{Total cost}) \\ = pc \int_{T_p}^{\infty} (T_N - T_p) dG(T_N) + (1-p) \left[ h \int_{T_p}^{\infty} (S - T_p) dF(S) \right. \\ \left. + g \int_{-\infty}^{T_p} (T_p - S) dF(S) \right]. \quad (2) \end{aligned}$$

It is easy to show that (2) is convex in the decision variable  $T_p$ . Thus, there exists a unique cost-minimizing starting point  $T_p^*$  at which the supplier should begin production of an order, which is characterized by the solution to the first-order condition

$$pcG(T_p^*) + (1-p)(g+h)F(T_p^*) = pc + (1-p)h, \quad (3)$$



Define the expected delay time at the optimal decision as

$$V(T_p^*) = E[LT - (RDD_N - T_p^*)]^+ = \int_{-\infty}^{T_p^*} (T_p^* - S) dF(S). \quad (4)$$

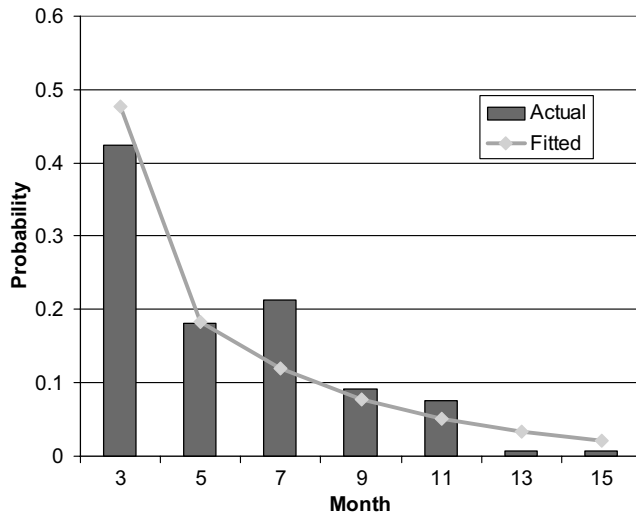
It is easy to show that  $V(T_p^*)$  is nonincreasing with the delay cost, nondecreasing with the cancellation cost, and nondecreasing with the holding cost, i.e.,  $\partial V(T_p^*)/\partial g \leq 0$ ,  $\partial V(T_p^*)/\partial c \geq 0$ , and  $\partial V(T_p^*)/\partial h \geq 0$ . The same is true for  $T_p^*$ , i.e.,  $\partial T_p^*/\partial g \leq 0$ ,  $\partial T_p^*/\partial c \geq 0$ , and  $\partial T_p^*/\partial h \geq 0$ .

#### 4.2. Functional Forms

To empirically reconstruct the cost parameters  $c$ ,  $h$ , and  $g$ , we need to make specific assumptions concerning the underlying distribution function for the arrival time of finalized information,  $T_N$ , and for the distribution function underlying  $S$ . For the arrival time of finalized ordering information,  $T_N$ , we assume an exponential distribution. Specifically, we assume  $G(x) = 1 - e^{-\alpha x}$ . Figure 4 compares the actual data we collected in the semiconductor equipment supply chain with an exponential distribution where  $\alpha = 0.21$ . A chi-square goodness-of-fit test cannot reject the null hypothesis that the underlying distribution is exponential.

Next, we need to specify a distribution function for  $S = RDD_N - LT$ . Testing separately, we find both

Figure 4 Distribution of  $G(\cdot)$



$RDD_N$  and  $LT$  are normally distributed, so a natural assumption would be that  $S = RDD_N - LT$  is normally distributed. However, the normality assumption makes the model less tractable and, therefore, in estimation problems similar to ours, a Weibull(2,  $\beta$ ) distribution is used instead. This distribution is also called the Rayleigh distribution.

In our context it is possible for  $RDD_N - LT$  to take on negative values, so we shift the Rayleigh distribution to the right by a constant  $\delta$ . The exact value of this shift  $\delta$  will be estimated jointly with the three cost parameters defined above. While the resulting distribution of  $S$  no longer is a Weibull distribution, it still has a relatively simple cumulative distribution function, with  $F(S) = 1 - e^{-\beta^2(S+\delta)^2}$ .

Substituting the functional forms into Equation (3) we obtain a simplified first-order condition:

$$pc e^{-\alpha T_p} + (1-p)(g+h)e^{-\beta^2(T_p+\delta)^2} = (1-p)g. \quad (5)$$

#### 4.3. Optimality Condition and Taylor Approximation

To generate a closed-form solution, we linearize the objective function in the neighborhood of a target starting date. We show in the appendix that Taylor approximation provides a reasonably good approximation and the solution obtained from the approximation is close to the true optimum value. Let the target starting time be  $\tau$ , around which we expand the exponential function in the above equation (below we set  $\tau$  to be three months in our estimation). That is,  $e^{-\alpha T_p} = e^{-\alpha \tau} - \alpha e^{-\alpha \tau} (T_p - \tau) + r_G$ , and  $e^{-\beta^2(T_p+\delta)^2} = e^{-\beta^2(\tau+\delta)^2} - 2\beta^2(\tau+\delta)e^{-\beta^2(\tau+\delta)^2} (T_p - \tau) + r_F$ , where  $r_G$  and  $r_F$  are the residual terms for those two expressions. The resulting first-order equation is

$$pc[e^{-\alpha \tau} - \alpha e^{-\alpha \tau} (T_p - \tau) + r_G] + (1-p)(g+h) \times [e^{-\beta^2(\tau+\delta)^2} - 2\beta^2(\tau+\delta)e^{-\beta^2(\tau+\delta)^2} (T_p - \tau) + r_F] = (1-p)g. \quad (6)$$

For ease of presentation, define  $\theta_1 = pc$ ,  $\theta_2 = (1-p)(g+h)$ , and  $\theta_3 = (1-p)g$ . Note that  $g$ ,  $h$ , and  $c$ , our key parameters of interest, can be fully recovered from  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ :  $g = \theta_3/(1-p)$ ;  $h = (\theta_2 - \theta_3)/(1-p)$ ;  $c = \theta_1/p$ .

Re-arranging terms, we obtain the following closed-form solution for  $T_p^*$ :

$$T_p^* = [\theta_1 e^{-\alpha\tau} + \theta_1 \alpha e^{-\alpha\tau} \tau + \theta_2 e^{-\beta^2(\tau+\delta)^2} + 2\theta_2 \tau \beta^2(\tau+\delta) e^{-\beta^2(\tau+\delta)^2} - \theta_3] \times [\theta_1 \alpha e^{-\alpha\tau} + 2\theta_2 \beta^2(\tau+\delta) e^{-\beta^2(\tau+\delta)^2}]^{-1}. \quad (7)$$

Our approximation approach has a direct analogy in the order-fulfillment process of the supplier under study. The supplier has a predefined milestone at which she initiates the manufacturing process. This milestone is then adjusted to reflect order-specific considerations, technical tool characteristics, shipment destination, or tool function. This is a natural analogy to Taylor series expansion, where an objective function can be approximated in the neighborhood of a given point, and actions can be adjusted in relation to that given point to improve the objective.

## 5. Estimation

We observed the semiconductor equipment supply chain at the interface of the buyer and the supplier for a total of 18 months, beginning in January 1999. For this time period, we created a complete history of forecast sharing via direct access to the online collaboration system described above. The buyer of the supply chain updates the online collaboration system on a monthly basis, providing the latest forecasts in the form of soft orders.

The system enables us to follow every individual equipment order from its initiation as a soft order to the completion of the installation and qualification procedure. For every order, we collected information on its total order lead time (from the time an order enters the system to the time it is fulfilled), the process technology generation that the equipment was designed to support, the function of the equipment, and the destination (geographic location of the fab) to which the equipment is to be delivered. Because we collected the data over an extended period of time, we were also able to observe information about order cancellations and requested changes to the delivery date. This database provided an accurate description of forecast sharing between buyer and supplier. The total number of soft orders in the sample was 143, of which 100 were firm orders.

### 5.1. Estimation Procedure

Assume that  $\alpha$  and  $\beta$ , the parameters for  $G(\cdot)$  and  $F(\cdot)$ , respectively, vary across orders, and that this variation can be explained by a set of independent variables. Let the observations corresponding to the individual pieces of equipment be indexed as  $i = 1, 2, \dots, I$ , and explanatory variables be indexed as  $j = 1, 2, \dots, J$ . Let any given  $\alpha_i > 0$  consist of a "base rate"  $\gamma_0$ , and an order-specific term that can be explained by a set of values of  $x_{ij}$ ,  $j = 1, \dots, J$ . In other words,  $\alpha_i$  can be written as

$$\alpha_i = \exp(\gamma_0 + \gamma_1 x_{i1} + \dots + \gamma_J x_{iJ}). \quad (8)$$

The approach of using covariates to endogenize a distribution parameter is common in empirical marketing models as well as operations research models. See Duenyas (1995) for an example of a customer-arrival rate that is explained through various attributes of the customer. Similarly, we define  $\beta_i > 0$  as

$$\beta_i = \exp(\rho_0 + \rho_1 x_{i1} + \dots + \rho_J x_{iJ}). \quad (9)$$

Our notation can be further simplified by labelling the explanatory variables in vector form as  $X_i = (1, x_{i1}, \dots, x_{iK})'$ , a  $J \times 1$  vector:

$$\begin{aligned} \alpha_i &= \exp(\gamma X_i), \\ \beta_i &= \exp(\rho X_i), \end{aligned} \quad (10)$$

where  $\gamma$  and  $\rho$  are the respective parameter vectors of dimension  $1 \times J$  to be estimated.

A complication arises in our empirical estimation. We did not have data on  $T_{p,i}$  (the supplier involved in this study did not keep the data). Consequently, we need to estimate the model using finishing time as follows. Denote  $FT_i$  to be the finishing time of order  $i$  (on which we had data), and let  $LT_i$  be the manufacturing lead time of order  $i$  (on which we did not have data). We have the following identity:

$$FT_i = T_{p,i} + LT_i. \quad (11)$$

That is, finishing time equals to starting time plus lead time. Let  $Y_i$  indicate a set of variables that influence manufacturing lead times. Then we can predict  $LT_i$  based on the following regression model:

$$LT_i = \eta Y_i, \quad (12)$$

where  $\eta$  is the parameter vector to be estimated.<sup>2</sup> To control for potential change in capacity utilization over time, we include a dummy variable for each half-year in the regression model. We also estimated our model with quarterly and bimonthly dummies and found the resulting estimates to be robust.

We assume that the finishing time  $FT$  is normally distributed:

$$FT_i \sim N(\mu_i, \sigma^2). \quad (13)$$

The normality assumption is common in regression analysis. To validate this assumption, we formally test this assumption in §7 (Model validation), where we find that normality assumption cannot be rejected. Using our previous results, we can predict the expected finishing time  $\mu_i$  as follows:

$$\begin{aligned} \mu_i = & \left[ \theta_1 e^{-\exp(\gamma X_i)\tau} + \theta_1 \exp(\gamma X_i) e^{-\exp(\gamma X_i)\tau} \right. \\ & \left. + \theta_2 e^{-\exp^2(\rho X_i)(\tau+\delta)^2} \right] \cdot \left[ \theta_1 \exp(\gamma X_i) e^{-\exp(\gamma X_i)\tau} \right. \\ & \left. + 2\theta_2 \exp^2(\rho X_i)(\tau+\delta) e^{-\exp^2(\rho X_i)(\tau+\delta)^2} \right]^{-1} \\ & + \left[ 2\theta_2 \tau \exp^2(\rho X_i)(\tau+\delta) e^{-\exp^2(\rho X_i)(\tau+\delta)^2} - \theta_3 \right] \\ & \times \left[ \theta_1 \exp(\gamma X_i) e^{-\exp(\gamma X_i)\tau} \right. \\ & \left. + 2\theta_2 \exp^2(\rho X_i)(\tau+\delta) e^{-\exp^2(\rho X_i)(\tau+\delta)^2} \right]^{-1} \\ & + \eta Y_i. \end{aligned} \quad (14)$$

When estimating finishing times, one cannot assume that two subsequent finishing times are independent of each other. Because of potential capacity constraints and/or congestion effects, a longer-than-average finishing time of the  $n$ th order—everything else equal—would increase the likelihood of a longer than average finishing time of the  $(n+1)$ st order. We therefore need to consider a joint distribution between  $FT_i$  and  $FT_{i+1}$ . Given the normality assumption of the finishing times  $FT$ , discussed above, we use a bivariate normal distribution (BVN) and specify  $(FT_i, FT_{i+1}) \sim BVN(\mu_i, \mu_{i+1}, \sigma^2, \sigma^2, \lambda)$ , where  $\lambda$  is the correlation coefficient. It follows that

$$FT_{i+1} | FT_i \sim N(\mu'_{i+1}, \sigma^2), \quad (15)$$

<sup>2</sup> We fitted quadratic, cubic, and log specifications to our model, but none of them improved the model fit significantly. Therefore, we keep the linear specification in this paper.

where

$$\mu'_{i+1} = \mu_{i+1} + \lambda(FT_i - \mu_i), \quad \sigma'^2 = \sigma^2(1 - \lambda^2). \quad (16)$$

Therefore, we can obtain the joint likelihood function:

$$\begin{aligned} L &= f(FT_1, FT_2, \dots, FT_n) \\ &= f(FT_1) f(FT_2 | FT_1) \dots f(FT_n | FT_{n-1}), \end{aligned} \quad (17)$$

where  $f(FT_{i+1} | FT_i)$  is the density to (15).

Maximum likelihood estimation (MLE) method can now be readily applied to obtain parameter estimates.

## 6. Estimation Results

We use four explanatory variables to predict the dependent lead-time variable  $LT$  (measured in months): tool generation, tool functionality, tool destination, and forecast changes. First, tool generation is a binary variable, which is set equal to one if the tool is a newly introduced technology. At the time of our study, one tool type was a new generation, and—based on our interviews—we expected this tool generation to have a longer lead time. Second, some tools in our sample are CMP (chemical mechanical planarization) tools, which are considered premium tools in the industry, and our interviews suggested that these tools would have longer lead times. We define a binary variable PREMIUM that takes on a value of 1 if the tool is of such function, and zero otherwise. Third, we expect tools ordered from and delivered to development fabs to exhibit a shorter lead time, so we include this binary variable DEVFAB to capture this effect. Fourth, our interviews indicated that during the delivery of production tools, changes with respect to RDD were of much higher importance than changes in equipment specifications,<sup>3</sup> therefore we include CHRDD as one key explanatory variable. Extent and frequency of forecast change in our data set are highly correlated (with correlation at about 60%). They both measure the degree of volatility of

<sup>3</sup> We did not consider changes in product specification because they are extremely rare in our context. The buyer follows a strategy in which every piece of production equipment has to be exactly the same, thus potential changes in specifications impact both new and existing tools.

**Table 1 Correlation Matrix**

	DEVFAB	NEW	PREMIUM	CHRDD
DEVFAB	1.000	0.247	0.096	-0.165
NEW	0.247	1.000	-0.074	-0.175
PREMIUM	0.096	-0.074	1.000	0.150
CHRDD	-0.165	-0.175	0.150	1.000

forecast changes. For this reason, we only focus on the frequency (CHRDD).

Finally, we control for the potential effect of capacity constraints at the supplier, and their impact on manufacturing lead time, by including fixed effects into our model. For example, if capacity was constrained during the first half-year of 2000, then a dummy variable for that half year would pick up the capacity effect in our sample. Note also that the correlation coefficient  $\lambda$  between two successive orders captures potential congestion effects at the supplier's operation.

The explanatory variables for  $\alpha$  and  $\beta$  are similar to the ones introduced in the regression on  $LT$ , namely, process generation (NEW), tool function (PREMIUM), and tool destination (DEVFAB). The correlation matrix of explanatory variables is reported in Table 1.

Like any other empirical study, our model can omit other crucial variables. To partly address this problem, we have conducted extensive interviews with both buyer and supplier, attempting to identify "qualitative" factors that we have not included in the analysis. One of the authors was on site for several months, while the other authors had weekly telephone meetings and had made several visits to various factory locations involved in this research. We observed that several crucial variables that could influence lead time, such as leadership in buyer and supplier organizations, task division within both organizations, and the relationship between the dyad was kept constant over the course of our study.

As MLE is a nonlinear method, results can sometimes be sensitive to starting points.<sup>4</sup> To ensure the

<sup>4</sup> We use GAUSS, a matrix programming language, to perform the estimation task. The algorithm for maximum likelihood estimation is BFGS.

**Table 2 Parameter Estimation Results: Full Model**

	1-step estimation		
	Estimate	Std. err.	$p$ -value
Imputed cost			
$g$	1.000	N/A	N/A
$h$	3.031	0.013	0.002
$c$	2.108	0.003	0.010
Lead-time shift $\delta$	0.068	0.017	0.000
Standard deviation $\sigma$	3.000	0.003	0.000
$T_p$			
Constant $\gamma_1$	-1.085	0.005	0.000
Develop $\gamma_2$	-4.326	0.005	0.000
New $\gamma_3$	-0.388	0.003	0.000
Premium $\gamma_4$	-9.721	0.003	0.000
Constant $\rho_1$	0.827	0.004	0.000
Develop $\rho_2$	-2.427	0.004	0.000
New $\rho_3$	-0.098	0.003	0.000
Premium $\rho_4$	-2.284	0.004	0.000
LT estimation			
Constant $\eta_1$	4.945	0.004	0.000
CHRDD $\eta_2$	0.966	0.004	0.000
Develop $\eta_3$	-3.024	0.003	0.000
New $\eta_4$	0.691	0.004	0.000
Premium $\eta_4$	-0.435	0.003	0.000
1HALF98 $\eta_5$	1.978	0.003	0.000
2HALF98 $\eta_6$	1.545	0.003	0.000
1HALF99 $\eta_7$	1.212	0.003	0.000
Correlation coefficient $\lambda$	0.207	0.003	0.000
Log likelihood		-227.211	

parameter estimates are not only locally optimal, we explored the parameter space by testing various sets of starting values. Specifically, we randomly generate 12,000 sets of starting values with each of the parameters drawn from a normal distribution with mean 0 and standard deviation 3, which covers the typical range of the parameter values from our observation.  $\theta_3$  is fixed at 0.700 (or equivalently  $g = 1.000$ ) for identification. We also restrict  $g$ ,  $h$ , and  $c$  to be non-negative. Table 2 reports the parameter estimates that yielded the globally maximum likelihood.

Converting estimated values for  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  into cost parameters  $g$ ,  $h$ , and  $c$ , we obtain the following values:  $g = 1.000$ ,  $h = 3.031$ ,  $c = 2.108$ . These results suggest that the manufacturer weighs the holding and

**Table 3** Nested Models

Model	Log-likelihood	$\chi^2$ -test statistic	p-value
Full model	-227.211		
Nested 1: DEVFAB	-232.952	11.482	0.003
Nested 2: NEW	-234.230	14.038	0.001
Nested 3: PREMIUM	-233.347	12.272	0.002

cancellation costs about three and two times higher than the cost of delay.

To test the effect of explanatory variables on the model, we test the following sequence of nested models:

1. Tool destination: DEVFAB. The corresponding hypothesis is  $H_{0,1}$ :  $\gamma_2 = 0$ , and  $\rho_2 = 0$ .
2. Tool process generation: NEW. The corresponding hypothesis is  $H_{0,2}$ :  $\gamma_3 = 0$ , and  $\rho_3 = 0$ .
3. Tool type: PREMIUM. The corresponding hypothesis is  $H_{0,3}$ :  $\gamma_4 = 0$ , and  $\rho_4 = 0$ .

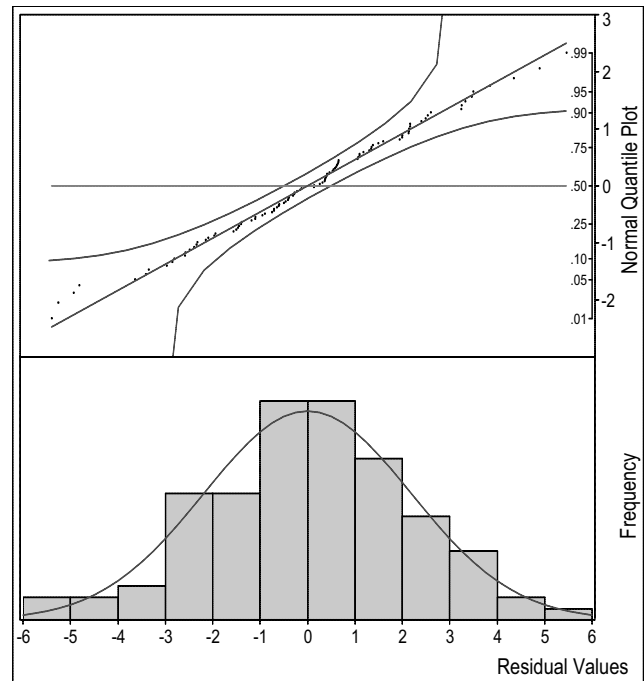
The results for those hypotheses tests are reported in Table 3. Likelihood ratio tests show that each of the hypotheses is rejected (see  $\chi^2$  test statistic in each table). Thus, all the variables indeed have predictive power in explaining lead time.

The parameter values of the dummy variables associated with time show that the capacity constraint does vary over time. The first half-year of 1999 appears to have the most constrained capacity. However the maximal difference in lead time across these fixed effects is 0.7 months (the average manufacturing lead time is 4.5 months), so we conclude that the capacity constraint affects the lead time, but not substantially. This is also consistent with the modest estimate on the correlation coefficient of finishing time. In-depth interviews with the managers at the supplier provide a plausible explanation. Apparently, orders from the buyer always receive a priority in production scheduling because he is the largest customer of the supplier. In fact, a special unit of marketing support was set up within the supplier organization specifically to facilitate the order-fulfillment process of the buyer.

## 7. Model Validation

We first validate our normality assumption. We perform both the Kolmogorov-Smirnov goodness-of-fit

**Figure 5** Test for Normality



test and the Chi-square test as described in Law and Kelton (2000). In both cases, the normality hypothesis could not be rejected. We also draw the Q-Q plot and distribution plot of residuals (Figure 5), providing also a graphical support for the normality assumption.

To further validate our model, we compare the finish time  $\widehat{FT}$  as predicted by our model, with actual finish time  $FT$ . Based on the estimates  $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\gamma}, \hat{\rho}, \hat{\delta}$ , and  $\hat{\eta}$ , we obtain a predicted value of  $FT$ :

$$\begin{aligned}
 \widehat{FT}_i &= \widehat{T}_{p_i} + \widehat{LT} \\
 &= [\hat{\theta}_1 e^{-\exp(\hat{\gamma}X_i)\tau} + \hat{\theta}_1 \exp(\hat{\gamma}X_i) e^{-\exp(\hat{\gamma}X_i)\tau} \\
 &\quad + \hat{\theta}_2 e^{-\exp^2(\hat{\rho}X_i)(\tau+\hat{\delta})^2}] \\
 &\quad \times [\hat{\theta}_1 \exp(\hat{\gamma}X_i) e^{-\exp(\hat{\gamma}X_i)\tau} \\
 &\quad + 2\hat{\theta}_2 \exp^2(\hat{\rho}X_i)^2 (\tau+\hat{\delta}) e^{-\exp^2(\hat{\rho}X_i)(\tau+\hat{\delta})^2}]^{-1} \\
 &\quad + [2\hat{\theta}_2 \tau \exp^2(\hat{\rho}X_i)(\tau+\hat{\delta}) e^{-\exp^2(\hat{\rho}X_i)(\tau+\hat{\delta})^2} - \hat{\theta}_3] \\
 &\quad \times [\hat{\theta}_1 \exp(\hat{\gamma}X_i) e^{-\exp(\hat{\gamma}X_i)\tau} \\
 &\quad + 2\hat{\theta}_2 \exp^2(\hat{\rho}X_i)(\tau+\hat{\delta}) e^{-\exp^2(\hat{\rho}X_i)(\tau+\hat{\delta})^2}]^{-1} \\
 &\quad + \hat{\eta} Y_i.
 \end{aligned} \tag{18}$$

**Figure 6** Model Validation

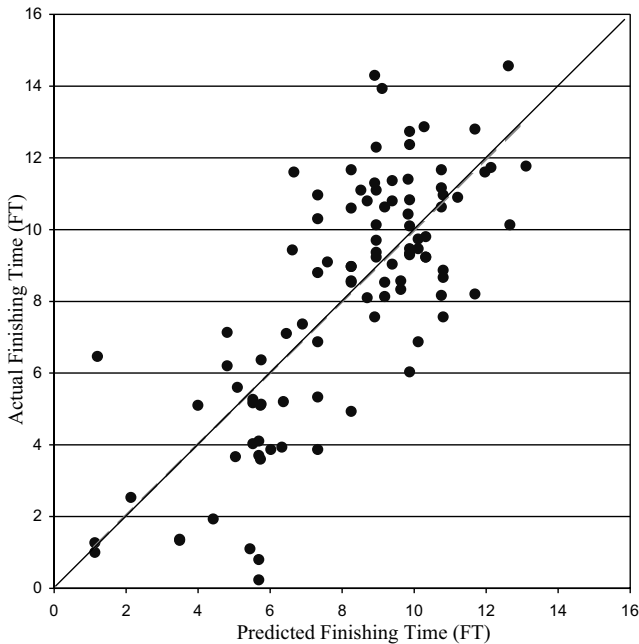


Figure 6 compares actual values  $FT$  and predicted values  $\widehat{FT}$ . The graph shows an overall good fit with no apparent systemic deviation from the identity line. An ordinary least square (OLS) regression of  $FT$  against  $\widehat{FT}$  confirms this observation:

$$FT = 0.000 + 0.999*\widehat{FT} \quad (19)$$

\*: significant at 1% level.

$R^2 = 40\%$ ; number of observations = 100.

$F$ -statistic: 137.97.

The results show an intercept that is statistically not different from zero ( $p$  value = 1.00), and a positive and significant slope ( $p$  value = 0.00). We conduct a test to check whether the slope is different from 1, and we cannot reject the hypothesis that slope is one ( $p$  value = 0.63). We also use White's general test (Greene 1993, p. 550) to test heteroskedasticity and could not reject the hypothesis that residuals are homoskedastic.

## 8. Policy Scenario Simulation

Our mathematical model, outlined in §4, combined with the parameter estimates for cancellation cost,

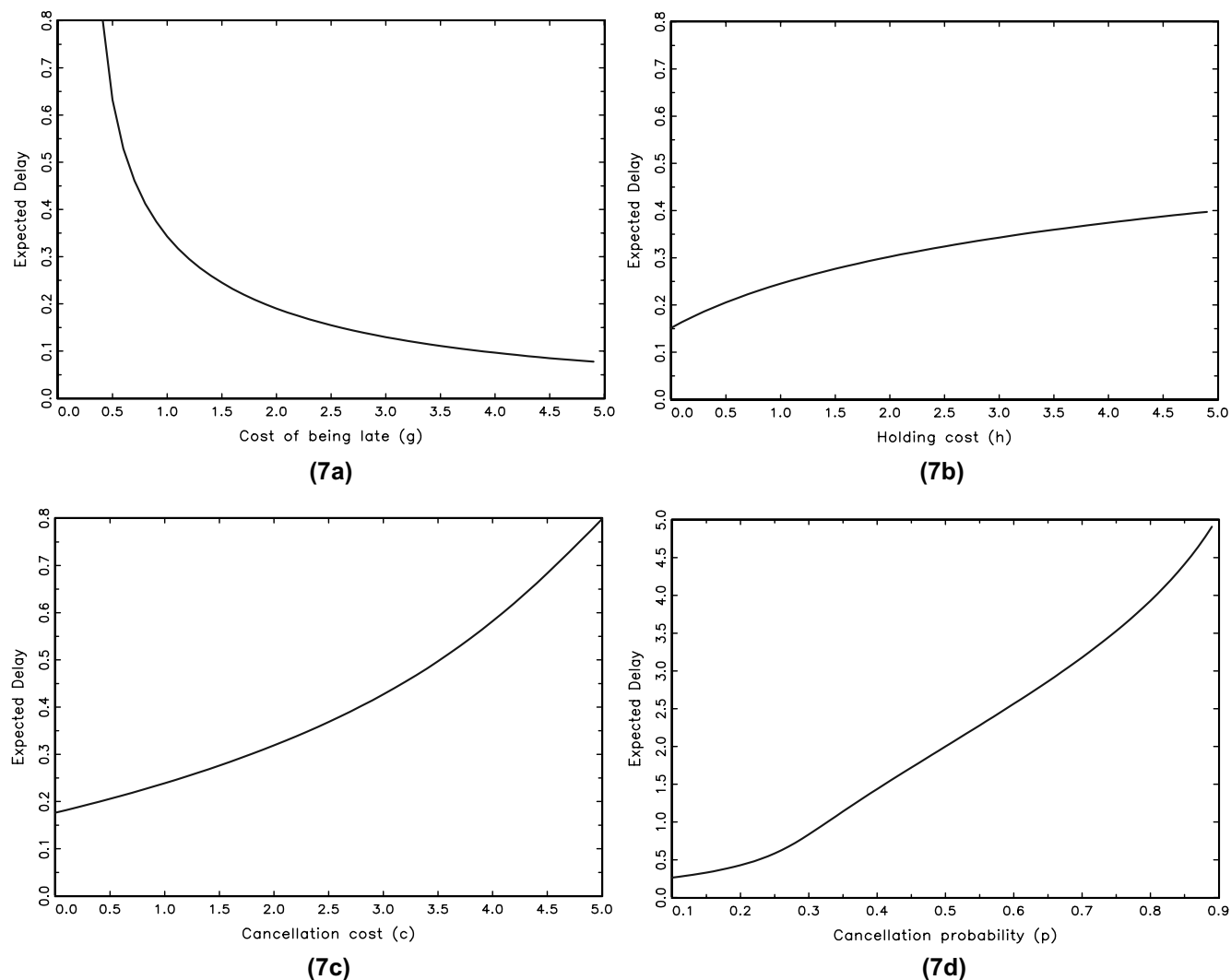
holding cost, and delay cost, enable us to analyze how modifications of cost parameters would impact the supplier's expected delivery performance. Note that the following analysis has been performed based on the cost parameters we obtained for the supply chain dyad underlying the present research study. While this methodology is generalizable beyond this setting, the magnitude of the following effects are likely to vary across different supply chain settings.

The buyer in our study was interested in the question of what would be the impact of a financial late shipment fee on the timeliness of deliveries. Economic intuition suggests that such penalty would increase the late shipment cost for the supplier, thereby encouraging her to commence production earlier (smaller  $T_p^*$ ). However, we can go one step further. Based on our analytical results and the empirical data, we can recompute the expected shipping delay with any given late-shipment cost parameter. This is depicted by Figure 7a, which shows the relationship between the cost parameter  $g$  and the expected slippage, as defined in Equation (4), averaged over all orders and expressed in months. Currently,  $g = 1$ , and the corresponding slippage is a little less than 0.4 months. Increasing late-shipment cost from  $g = 1$  to  $g = 2$  translates into a 0.15-month reduction in late shipment.

Next, consider the impact of holding cost  $h$  on the expected shipping delay (Figure 7b). Again, the status quo corresponds to  $h = 3$  and a delay of 0.34 months. Now consider what happens if holding costs are cut in half. For example, the buyer could accept the equipment prior to the specified requested dock date (and of course, also pay for it earlier). Alternatively, the buyer could partially reimburse the supplier for the holding costs. Reducing holding costs from 3 to 1 would lead to a 0.1-month reduction in shipping delay. While obviously the financial burden of capital cost now rests with the buyer (who pays for the equipment and then leaves it idle up to the time of actual need), the buyer might still be better off because shipment delays can put the production of entire fabs at risk, thereby having a much larger impact on the bottom line.

Similarly, Figures 7c and 7d investigate the relationship between changes in cancellation cost  $c$  and

**Figure 7** Impact of Delay Cost  $g$  (7a), Holding Cost  $h$  (7b), Cancellation Cost  $c$  (7c), and Probability  $p$  (7d) on Expected Delay

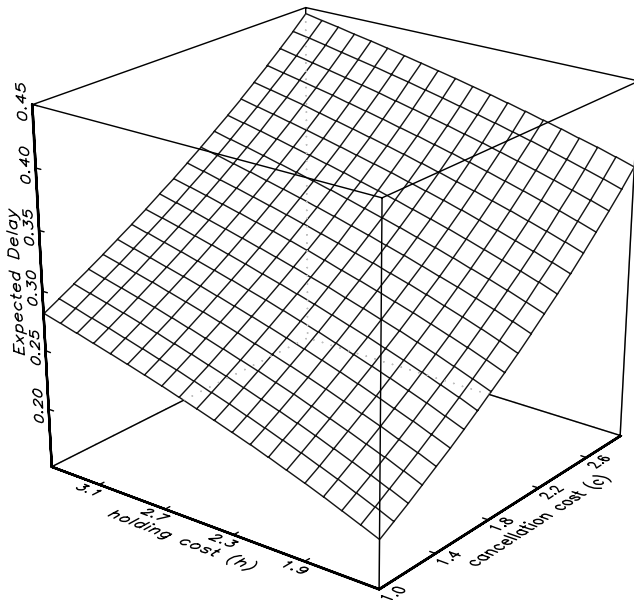


cancellation probability  $p$ . A reduction in cancellation cost could be achieved if the buyer would take over some of the cost incurred by the seller in the case of cancellation (e.g., procurement cost). Alternatively, the buyer could develop specifications with more standardized components in it, which would allow the supplier to reuse entire subassemblies for another customer after receiving notice of cancellation.

Interestingly, none of the cost changes outlined above is able to eliminate the expected delay. The reason for this lies in the complex trade-off that the supplier faces when deciding on the optimal time to start working on an order  $T_p^*$ . This trade-

off not only involves two forces but rests on a subtle balance between three forces. Hence, even a large improvement along one dimension will lead to only a small change in the supplier's decision—and thereby the expected delay—as the other two forces are still unchanged. Consequently, substantial changes in expected delay can only be achieved by changing at least two of the cost parameters jointly (opposed to changing them one at a time). This is illustrated in Figure 8. We see that reductions in holding cost and in cancellation costs actually complement each other, as opposed to acting as substitutes. A 50% reduction in both of them ( $h = 1.5, c = 1$ ) would reduce

**Figure 8** Joint Impact of Cancellation Cost  $c$  and Holding Cost  $h$  on Expected Delay



the expected delay by 0.15 months, while the two changes implemented individually would lead to a lower reduction of the delay (0.1 months for the reduction in  $h$  and 0.03 for the reduction in  $c$ ). That is, the joint changes are superadditive.

## 9. Discussion and Conclusion

Our results indicate that the supplier fears holding costs and order cancellations, making her averse to commencing order fulfillment based on soft orders. This results from the fact that the supplier's effort, including procurement of components and the actual building of the equipment, is very customer specific. We also find that the supplier perceives holding cost and cancellation cost as much more important relative to the cost of delay.

The large emphasis on early-completion cost relative to late-completion cost clearly does not mirror the overall cost for the supply chain. If the tool is finished early, it remains at the supplier's plant, and only traditional inventory holding costs are incurred. However, a late shipment of the tool can lead to idle time and lost output at the buyer's fab, which is associated with substantial margin losses, which

are magnitudes larger than the holding cost for a piece of equipment. This suggests there is a lack of coordination in the supply chain that can lead to sub-optimal performance.

The current situation can be partly explained by the fact that the customer in this study is a dominant player in semiconductor products and is the single largest customer of the supplier. This creates an imbalance of power in the buyer-supplier relationship, which is why the customer can provide forecasts without commitment and can change order forecasts without penalty, leaving the supplier to bear all the resulting costs.

The overall supply chain performance could be improved if the customer were willing to share some of the holding cost. One operational way of doing this would be if the customer accepted the tool delivery for some time window prior to the RDD. This would reduce the expected holding cost for the supplier and thereby move the optimal starting point  $T_p^*$  forward in time. In general, given the high degree of customization demanded by the buyer, supply chain performance also could potentially be improved if the buyer were ready to share some of the risk of cancellation. This would have two beneficial effects. First, it would reduce the supplier's cancellation cost, moving the optimal starting point  $T_p^*$  forward in time. Second, it would make the forecast more credible, and thereby rebuild some of the trust missing in the system. In the presence of a cancellation fee, phantom orders become costly to the buyer, allowing the supplier to have more confidence in the soft orders. Currently, forecasts are provided by the individual fabs of the customer. While all fabs are part of the same company, they also have local objectives. In such a setting, those fabs do not incur the full negative externalities that cancellation has on the reputation of the company as a whole. This suggests that a better coordination of forecasting activities across fabs could be beneficial.

The sharing of demand forecast information does not have a positive value for the supplier, who distrusts these data and delays the production start for the equipment. In our situation of voluntary compliance (the buyer cannot monitor the supplier), financial incentives are needed to make signals related to



the forecast credible (Cachon and Lariviere 2001). As the buyer, who is in control over the design of the coordinating mechanisms in the supply chain, does not incur any financial loss in case of a cancellation, such credible signaling is not possible. This would change if the buyer were to pay some cancellation fee (potentially as a function of time). While having the right of free cancellation is obviously attractive to the buyer and potentially saves him some direct out-of-pocket cost, the buyer pays a (much higher) price indirectly, resulting from shipment delays and long tool delivery lead times generated by the supplier's response to the current system structure.

We believe that these results are of substantial interest, both from an academic and a direct managerial perspective. On the academic side, these results provide the first econometric evidence of problems related to forecast sharing. While there is a rapidly growing stream of research following Lee et al. (1997), no previous study could econometrically demonstrate the existence of the coordination problems. While the magnitude of the parameter estimates we report in this study is specific to our research setting, our method can easily be implemented to obtain estimates in other supply chain dyads.

From a managerial perspective, our results demonstrate that information sharing by itself is not sufficient to build superior supply chain performance. Our results were presented to senior executives at both the customer and the supplier. In response to our study, the customer started several projects with the objective to overcome some of the credibility problems. For example, one project that the customer initiated attempts to acknowledge the uncertainty inherent in forecasts by communicating it explicitly via a range of possible orders (an interval) as opposed to ignoring it (simply sharing point-based forecasts). In the context of product development, such set-based information exchange has been documented to lead to improved performance.

Finally, we believe that our work not only serves as the empirical foundation for much of the contracting research, but that it also provides a fruitful starting point for future research. A larger empirical study could analyze how the cost parameters that we estimated change over time. For example, one

would expect that the cancellation of an order directly increased the supplier's perception of cancellation cost in the subsequent period. Another interesting research opportunity relates to how the forecast is shared. Similar to the field of concurrent product development discussed in the literature review, where there has been a recent trend toward set-based—as opposed to point-based—information exchange, the buyer could provide multiple scenarios of demand to the supplier or could even share a confidence interval. This would be consistent with established supply chain concepts such as minimum-purchase commitments, and its effect on forecast credibility would be interesting to study both analytically and empirically.

### Acknowledgments

The authors thank the management teams of the supply chain dyad, who generously provided their internal data. They are also grateful for the constructive comments by the departmental editor, the associate editor, and three anonymous referees.

### Appendix. Taylor Expansion

The function we want to approximate is Equation (5), the first-order condition for the cost-minimization problem.

$$pce^{-\alpha T_p} + (1-p)(g+h)e^{-\beta^2(T_p+\delta)^2} = (1-p)g.$$

As before, denote

$$\begin{cases} \theta_1 = pc, \\ \theta_2 = (1-p)(g+h), \\ \theta_3 = (1-p)g. \end{cases}$$

We can rewrite the first-order condition in the form of  $f(x) = 0$ , i.e.,

$$f(x) = \theta_1 e^{-\alpha T_p} + \theta_2 e^{-\beta^2(T_p+\delta)^2} - \theta_3 = 0.$$

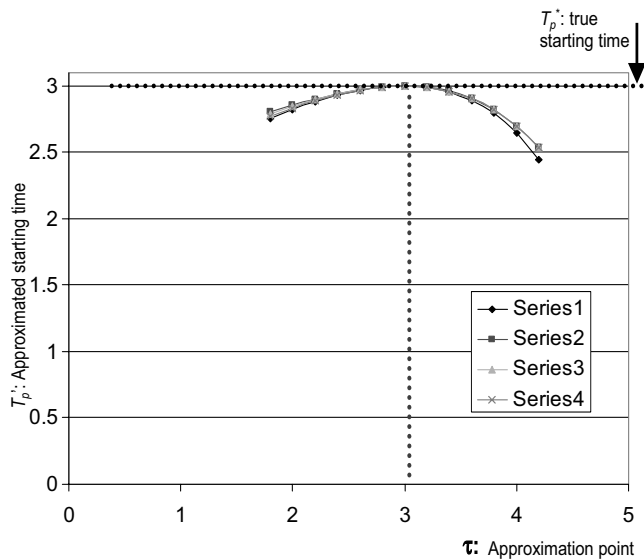
Note that  $f(x)$  is monotonically decreasing in  $T_p$ . Approximate  $f(x)$  around a given point  $\tau$  by the tangent line,  $f'(x)$ , passing through  $(\tau, f(\tau))$ . Solving  $f'(x) = 0$  will give us the approximated solution  $T_p'$ , where

$$\lim_{\tau \rightarrow T_p^*} T_p' = T_p^*.$$

Ideally, we want  $T_p'$  to be close to  $T_p^*$ . We conducted a numerical study to confirm this proximity for a wide set of parameter ranges, including cases where  $\tau$  differs significantly from  $T_p^*$ . Toward this goal, we first randomly sampled different parameter values and obtained  $T_p^*$  numerically. Next, we calculate the approximated solution  $T_p'$  for the given value of  $\tau$ .

The numerical study shows that for most of the parameter values,  $T_p'$  performs reasonably well. An example of this is illustrated

Figure 9 Taylor Approximation



in Figure 9. In the example, the true solution is  $T_p^*$  is 3. Varying  $\tau$  between 2 and 4 does not lead to big changes in  $T_p'$ , whose values remain in the interval  $[2.6, 3]$ . In general, we find that for the parameter settings in our study, varying  $\tau$  by  $\pm 30\%$  still yields approximated solutions,  $T_p'$ , that are within 10% of the optimal solution  $T_p^*$ . Thus, Taylor approximation indeed provides a good representation of the original problem.

## References

Biddle, B. F. M. 1998. Boeing to cut 747 output 30% in 1999 and to curtail production of its 777. *Wall Street J.* (June 10).  
 Cachon, G. P., M. A. Lariviere. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Sci.* 47(5) 629–646.

Chen, F. 2004. Information sharing and supply chain coordination. T. de Kok, S. Graves, eds. *Handbook of Operations Research and Management Science: Supply Chain Management*. North-Holland, Amsterdam, The Netherlands.  
 Cohen, M., J. Eliashberg, T.-H., Ho. 1996. New product development: The performance and time-to-market tradeoff. *Management Sci.* 42 173–186.  
 Cole, J. 1997a. Boeing pushing for record production, finds parts shortages, delivery delays. *Wall Street J.* (June 26).  
 —. 1997b. Boeing suppliers are beating the heat as jet maker pushes to boost output. *Wall Street J.* (September 16).  
 Duenyas, I. 1995. Single facility due date setting with multiple customer classes. *Management Sci.* 41(4) 608–619.  
*Flight International*. 2001. Airliner orders plummeting.  
 Greene, W. 1993. *Econometrics Analysis*. Macmillan Publishing Company, New York.  
 Kelly, K. 1995. Burned by busy signals. *Business Week* (March 6).  
 Krishnan, V., S. D. Eppinger, D. E. Whitney. 1997. A model-based framework to overlap product development activities. *Management Sci.* 43(4) 437–451.  
 Law, A. M., D. W. Kelton. 2000. *Simulation Modeling and Analysis*. McGraw-Hill, Inc., New York.  
 Lee, H. L., V. Padmanabhan, S. J. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Sci.* 43(4) 546–558.  
 Loch, C. H., C. Terwiesch. 1998. Communication and uncertainty in concurrent engineering. *Management Sci.* 44(8) 1032–1048.  
 Lucas, R. E. 1976. Econometric policy evaluation: A critique. K. Bruner, A. H. Metzler, eds. *The Phillips Curve and Labor Markets*. North-Holland, Amsterdam, The Netherlands.  
 Terwiesch, C., A. De Meyer, C. H. Loch. 2002. Exchanging preliminary information in concurrent engineering: Alternative coordination strategies. *Organ. Sci.* 13(4) 402–419.  
 Zarley, D. 1996. Backlogs plaque HP—Resellers place phantom orders to get more products. *Computer Reseller News* (May 6).

Accepted by Fangruo Chen, former department editor; received November 28, 2001. This paper was with the authors  $4\frac{1}{2}$  months for 2 revisions.