RC24260 (W0705-075) May 15, 2007
Mathematics

IBM Research Report

Demand Distortions and Capacity Allocation Policies

Jayanth Krishnan, Paul Kleindorfer
The Wharton School
University of Pennsylvania

Aliza Heching
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Demand Distortions and Capacity Allocation Policies[1]

Jayanth Krishnan
jayanth@wharton.upenn.edu

Paul Kleindorfer
kleindorfer@wharton.upenn.edu

Aliza Heching
ahechi@us.ibm.com

Updated May 4, 2007

Abstract

In this paper, we describe a study of the consequences of capacity allocation mechanisms at a large semiconductor manufacturer. We use a panel data set with 300,000 observations spanning five years of forecasting and order transactions between a supplier and customers at the supplier's manufacturing facility. Using Mixed Linear Models (MLM), we study the mutual interplay between a supplier's capacity allocation mechanism and customers' demand forecasts and orders. The results of our model suggest that the interaction causes two types of distortions: inter-temporal and cross-sectional. Temporal forecast distortions result in forecast churn (short-term volatility in forecasts), undesirable forecast smoothing (a.k.a., batching), and customers exiting the facility. Cross-sectional forecast distortion (contagion) is characterized by temporal forecast distortions spreading across individual customers' forecasts, attributable to the negative externalities imposed by commonly used capacity allocation policies. We present empirical tests for the presence and significance of these distortions in our data and note the impact of churn on buffer inventory stocks and contagion's effect on risk pooling efforts. Our findings have implications for contracting, capacity planning, and CPFR initiatives in supply chains.

Keywords : capacity allocation, churn, contagion, forecast distortions

[1] We would like to thank Alliance Center for International R&D (INSEAD), CIBER-Penn Lauder Dissertation Grant, Risk and Decision Processes Center (The Wharton School) for their generous financial support in conducting this project. The authors are especially grateful to Jim Feldhan, CEO of Semico Research Corporation, for his early support of this research and the industry contacts he provided us. We would also like to thank Vishal Gaur, Serguei Netessine, Marshall Fisher and conference participants at INFORMS-San Francisco, MSOM-Atlanta and INFORMS-Pittsburgh for their useful comments.

1. Introduction

This paper studies a problem endemic to capital-intensive industries, namely, that of planning and managing forecasts and orders from customers to a supplier. We seek to understand how a supplier's own capacity allocation policies can distort his view of downstream demand in a supply chain. We conduct our analysis using proprietary data from the years 1999-2004 from a large semiconductor firm. The issues examined in this study namely, forecast evolution, information sharing, contract design and capacity allocation policies, have ramifications for a broad spectrum of capital-intensive industries, including semiconductors.

The primary goal of this paper is to investigate the effects of capacity allocation on customer behavior, i.e., to determine how customers revise their future forecasts and orders when faced with capacity allocation. By capacity allocation we refer to situations where the total order quantity that a supplier receives from his customers exceeds his net available capacity and hence the supplier allocates his limited capacity to the customer base. We refer the reader to Cachon and Lariviere (1999a) for a review of some commonly used allocation policies such as priority-based and proportional contracts. To analyze capacity allocation policies, existing literature has focused on the dyadic relationship between a supplier and his customers. In this paper, we point to the presence of spillover effects of such dyadic interactions across customers. We will show that these spillover effects can, in fact, be quite significant in the presence of uncertain supply conditions and undisclosed allocation policies.

Three features of the semiconductor industry make it an interesting area for empirical study of supply-demand coordination and capacity allocation: (i) Semiconductor manufacturing is typically build-to-forecast, so that forecast churn and inaccuracies can lead both to capacity mis-utilization as well as tensions with customers. (ii) The semiconductor industry is highly capital intensive; current initial fab investment costs around $3 billion and (according to S&P Industry Survey, 2005) is increasing in real terms as technological progress in wafer fabrication, following Moore's Law (Intel Press Release, 2005), continues unabated. In light of such large capital investments, sales and operating groups within firms face intense pressure to maintain high asset (fab) utilization by attracting and retaining customers with large orders. Given this backdrop, capacity allocation and demand distortion can significantly impact profit. (iii) The semiconductor industry offers a wide range of products – from commodities products (e.g., DRAMs) to proprietary products (e.g., Intel's Centrino processor) to customized products that are designed for use by one customer on a specific application (e.g., Foundry and ASICs). This richness in product suite, together with the other features noted of semiconductor manufacturing, makes forecasting a center piece of effective capacity management.

The paper proceeds as follows. In section 2, we present the academic literature that guides our hypotheses, in section 3 we present industry practices and the relevant business issues, followed by a description of the data from a manufacturer in section 4. In section 5, we describe the forecasting behavior of customers that we observed in the data that led us to develop hypotheses about information distortions in the supply chain of this supplier in section 6. In section 7, we present an econometric model followed by discussion of the results from estimation of this econometric model in section 8. In section 9, we conclude with managerial implications of our findings and suggestions for future research.

2. Literature Survey

In this paper we investigate supply-demand mismatch problems and the relationship to these of customers' forecasting policies and suppliers' allocation policies. Four areas of the literature are directly pertinent to our paper: 1) forecasting models that buyers use to make ordering decisions, 2) value of information in forecast updates, 3) suppliers' allocation policies and 4) the interaction between the forecasting and allocation policies. Since our paper spans all these four areas, we can only review the literature in brief, focusing on foundational and recent papers.

Hausman (1969) alluded to the quasi-Markovian property of forecasts and the emergence of lognormal models (in finance theory) to approximate existing forecasts for use in dynamic programming models of production. Heath and Jackson (1994) generalized the Hausman model to a martingale Model of Forecast Evolution (MMFE) that can be used in a rolling horizon multi-period format. For our paper, we adopt the

MMFE convention for modeling forecasts and forecast updates. The Heath and Jackson model has since been adopted in prescriptive production planning models (Graves et al., 1998; Toktay and Wein, 2000). Cattani and Hausman (2000) discuss why forecast updates may not reflect the true underlying demand distribution. Miyaoka and Hausman (2004) analyze the design of inventory policies at both upstream and downstream supply chain partners to minimize bullwhip effects. Our work borrows from the modeling framework of Heath and Jackson (1994), but it differs from the above literature in its intent, i.e., we take an empirical approach to investigate information distortions rather than seeking prescriptive solutions to known forms of information distortion (such as the bullwhip effect).

In our paper, we use time series models in conjunction with MMFE notation to understand the value of private information signals that a supplier conveys to the buyer through his capacity allocation policy. Lee, So and Tang (2000) quantify the value of information sharing in a two-stage supply chain for a single-period lagged (i.e., *AR(1))* demand model with non-negative autocorrelation coefficient. Gaur et al. (2005) extend the above model to a general *ARMA(p,q)* setting and show that the value of information sharing depends on the time series structure of the underlying demand. Our work contributes to this stream of literature by modeling and estimating the effect that the supplier's allocation policy (supply information) has on the statistical structure and magnitude of demand forecasts over time.

The foundational papers in the area of capacity allocation are Cachon and Lariviere (1999a,b, 2001), which study incentive contracts to elicit truthful forecasts from downstream retailers (customers) who may inflate their forecasts in anticipation of future capacity allocation. They consider a setting in which a single supplier faces demands for its scarce capacity from multiple retailers/customers, where the latter face deterministic demands unknown to the supplier. When there is excess demand, the supplier allocates (rations) capacity. Cachon and Lariviere consider a broad class of allocation policies and demonstrate how each is prone to manipulation by downstream customers of inventory/capacity. They also show that there exists a class of mechanisms that induce truthful revelation of customer demands, but that do not maximize total retailer profits. The key idea that we seek to build on is that ***allocation policies can mask a customer's***

true demand. By extending the Cachon and Lariviere framework to a stochastic demand setting, we retain Cachon and Lariviere's key intuition on the incentive effects of allocation policies on customer forecasts. We obtain additional insights on the effects of individual forecast biases on aggregate, i.e. facility-level, demand from a group of customers facing possible capacity constraints.

There are several other notable contributions to the interactions between forecasting and capacity allocation. Kaminsky and Swaminathan (2001) develop a forecast refinement method to plan for production in a capacitated facility. Heching and King (2005) discuss supplier's inventory management policies for minimizing risks arising from long lead-times, a bane of the semiconductor industry. At the interface of customers' forecasting processes and supplier's allocation policies, there arises the crucial question, *how can better downstream forecast and information sharing practices improve upstream capacity planning?* Aviv (2001, 2002) discusses the merits of collaborative forecasting practices under different demand conditions. Ferguson (2001), and Ferguson et al. (2003) study the customer's problem of simultaneously managing demand forecasts and making commitments to the supplier. This framework of managing both supply and demand uncertainty is central to the ideas of our paper as well. Gavirneni et al. (1999) characterize the value of information sharing when capacity is costly and limited, while Milner and Kouvelis (2002) suggest that in managing both demand and supply uncertainty, there exist complementary benefits to improving demand forecasts and improving supplier flexibility. Plambeck and Taylor (2003) study the role of repeated buyer-supplier interactions in a strategic setting that has implications for the dynamic evolution of capacity planning models. We note that this stream of literature places capacity management as the center piece of attention in the forecasting-capacity interaction. In the present paper,we investigate the endogeneity between capacity allocation policies and forecast accuracy.

Related to the literature on supply allocation is that of bullwhip effect, attributed to information distortions in a supply chain (Lee, Padmanabhan and Whang, 1997). Empirical evidence in the OM literature concerning the nature of buyer-supplier interactions is of relatively recent vintage (Cachon et al., 2005; Ren et al., 2004). Cachon et al. (2005) investigate the prevalence of the bullwhip effect across industries and echelons of the supply chain and concludes that while it is evident at the wholesaler level, it is not evident at other echelons of the supply chain (and particularly not at the manufacturer level). In related experimental work, Croson and Donohue (2006) study the causes of the bullwhip effect in a controlled laboratory setting. Terwiesch et al. (2004) investigate forecast sharing between buyers and suppliers in the semiconductor equipment industry. They show how the past actions of a buyer affect the future actions of the supplier. A key differentiator between Terwiesch et al. (2004) and our paper is that we focus attention on the negative externalities that arise out of individually rational actions across buyers using a common supplier while their work is centered on the dyadic interaction between a buyer and supplier. Our work on negative externalities and group behavior is inspired by Schelling (1978) in his classic treatise on "Micromotives and Macrobehavior" and by the early work of Forrester (1958) on interaction effects in dynamic industrial environments.

To summarize, we study the effect of supplier's allocation policies on buyers' forecast (distortions). We borrow the notation of the MMFE model pioneered by Heath and Jackson (1994) to detect information distortions (such as churn, contagion and batching, all defined more precisely below, and the well-known bullwhip effect), that are group-level effects that arise out of supplier and buyer decisions and interactions in a capacitated (semiconductor) supply chain.

3. Context and Industry Practice

Semiconductor manufacturing complexity and leadtime are highly dependent on the degree of product customization. The manufacturing process begins with a raw wafer, ordinarily purchased from a third-party. Then, chemicals are applied to and removed from the wafer to create specific patterns. The manufacturing process may involve multiple iterations of this etching and patterning, resulting in multiple layers of the wafer. Total cycle time for wafer fabrication is typically 40-70 days; however, this process can take longer for more complicated products. Following wafer fabrication, the next step in semiconductor manufacturing is packaging. The wafer is cut into individual integrated circuits (die), the die are attached to the packaging, circuits are connected to the outside of the packaging, and final packaging steps are completed. Packaging typically takes 14 days. Various testing takes place throughout the manufacturing cycle and adds from 1-4 days to the cycle time.

Customized semiconductor devices typically have significant business integration costs associated with establishing a customer-supplier relationship. These include, for example, the costs of prototyping, developing masks (specification of the circuit design through layered masks), and testing and validation procedures. Thus, a customer often sources a semiconductor design from a single or very few suppliers. Downstream, there are additional steps in logistics and assembling of the products using semiconductors, from automobiles to consumer electronics. In this paper, we focus on the supplier and its immediate customers. The number of customers that could make demands over a two-three year period at a large semiconductor manufacturer (such as the one we study here) would be on the order of hundreds, with major buyers comprising only 10-20% of the total number of customers but accounting for 70-80 % of total volume.

3.1. Forecasting Process

In this section we describe the forecasting process (see Figure 1 for graphical illustration). The majority of the procurement contracts require the customer to provide the supplier with rolling horizon demand forecasts prior to placing actual orders. The forecast horizon begins twelve months prior to the request date, where the product request date is the date that the customer wishes to receive the product. The supplier uses the longer horizon demand forecasts to help with capacity planning; in some cases of anticipated capacity constraints, the supplier may contract for outside sources of capacity. Thus, accurate forecasts are advantageous both to the supplier and to the customer; the supplier can better coordinate demand and capacity (resulting in improved profitability) and the customer can enjoy better demand fulfillment. Contract terms also often limit the allowable percentage modification of month-over-month demand forecasts (see Tsay and Lovejoy, 1999); within the production leadtime contracts often contain terms specifying the conditions under which customers may change the magnitude and delivery dates of their orders. The purpose of contracting terms is to assure an orderly commitment of capacity to customers. From the supplier's perspective, such terms are intended to protect the supplier against the risk of customers radically changing or canceling demand quantities, especially in the case of custom products, once production has begun. In practice, however, such contract terms may be modified due to the competitive business environment in the semiconductor industry. Nonetheless, customers are encouraged by a variety of means, both formal and informal, to abide by the terms of contracts and the forecasting constraints imposed by these. However, not all customer orders are preceded by forecasts. Depending on the nature of the product, some customer orders arrive in a "spot-market" like fashion, and the corresponding orders may be served without any prior demand forecasts.

Following the sequence of demand forecasts the customer places a firm order. A firm order consists of an order by a given customer for a given part number (p/n) for a specified request quantity, denoted by Q_r, to be delivered on a given request date. The supplier may not be able to meet all demands as they ultimately materialize, and the supplier's sales and operations planning (S&OP) interface then makes the necessary adjustments, allocating capacity to the various customers and orders. The supplier responds to the request quantities with a quantity Q_c, called the "commit quantity". We note that when the supplier responds to the customer, he is making decisions in the face of uncertainty. Some demand is known with certainty (firm orders that the supplier has already received from customers) while some firm orders have not yet been received by the supplier though the supplier anticipates receiving them. Both customer and supplier face additional uncertainties due to the uncertainty of how many good die will yield from each wafer.

Our interests in this paper are to understand the mutual interplay between the customers' behavior in the face of capacity allocation and the supplier's reaction to customers' forecast updating process. Our analysis starts from a point where customers have already established a working relationship with the supplier and are otherwise qualified to make understandable demands on the capacity of the facility. These might range from simple procedures for commodity products, which could be made using off-the-shelf masks and procedures, to highly complex pre-testing and prototyping for customized products for which the manufacturer is the sole source due to the large set-up costs for prototyping.

Consider now the structure of forecasting (illustrated in Figure 2). There are two types of customers: those who forecast prior to making an order and those who place orders without prior forecasts. The forecasts occur in a 12-month rolling horizon format, i.e., the supplier begins to record forecasts twelve months ahead of the request date. Following the sequence of forecasts beginning up to 12 months in advance of actual order release date, the customer requests a quantity Q_r of a specific part number for order release and delivery after the elapsed in-plant manufacturing time. If the customer arrives in "spot-market like" fashion, he places a request Q_r, with no associated forecasts. The supplier responds with the commit quantity Q_c which is delivered to the customer on his requested date (or on a mutually agreed upon "commit date"). Customers who receive less than their request quantity (Q_r) may respond in various ways illustrated in Figure 2, from backordering to batching of future orders to seeking another supplier (i.e., exiting the facility).

4. Data Description

We analyze a proprietary panel data set with two cross-sectional components (customer and part number) and two longitudinal components (a series of date-specific quantity requests and the corresponding -- up to 12-month -- series of forecasts for each such quantity request). We assume that each customer's contract is devoid of any significant temporal variations over the 2000-2004 timeframe. This assumption is reasonable, as contracts in semiconductor manufacturing are typically long-term in nature. The contents of the database are summarized in Table 1. The following is a description of the variables contained in Table 1:

CUST = unique Customer ID. CUST represents a customer who either reported a forecast or placed an order with the supplier; the data set contains 2,565 unique CUST values.

P/N = unique part number for which the customer is reporting a demand forecast or placing an order. Our data set contains 11,868 unique P/N's. We note that products that undergo an engineering change or new release are typically assigned a new P/N. Thus, some of these 11,868 unique P/N's may represent similar products. The data set contained 28,800 unique CUST-P/N combinations suggesting that an average customer dealt in approximately 11 unique part numbers.

SPOT = binary variable indicating whether a unit of the cross-section (CUST-P/N) did (SPOT=0) or did not (SPOT=1) provide demand forecasts.

SYSEXIT= binary variable associated with each CUST-P/N pair. SYSEXIT=1 indicates that the CUST-P/N left the supplier with no future forecasts or orders. In this data set, 20,523 of the CUST-P/N combinations (approximately 70%) have SYSEXIT=1. This annual turnover rate is approximately 75% at both the CUST level and P/N level. This phenomenon is not necessarily indicative of customer dissatisfaction. Alternative explanations include, for example: (i) 30-35% of P/N's experience turnover due to engineering changes. Each time a product undergoes an engineering change, it is assigned a new P/N (thereby eliminating the original customer-p/n combination); (ii) Products under prototype are assigned a P/N. When they become mature products, they are assigned a new P/N; (iii) Before placing the final order, a customer may transfer the order to a subcontractor, thereby changing the customer on the order (and eliminating the original CUST-P/N pair); (iv) Natural end of a product lifecycle. Products such as mobile phones have an average 12-month lifespan. Then, the part number is eliminated.

START = first month that demand forecasts (for a given CUST–P/N pair) were received relative to the request date (for the given pair). In a 12-month forecasting horizon, START can take on values in the range [-12, 0] where START=-12 represents a demand forecast for demand 12 months prior to the request date. Similarly, START =3 represents a demand forecast for demand three months prior to the request date, etc.

FINISH = month that the final forecast (for a given CUST–P/N pair) was received relative to the request date. FINISH can take on values in the range [-12, 0]. We observe in the data that the typical START value is -7 and the typical FINISH value is -2, however, the standard deviations around these values are significant (approximately 3-4 months) suggesting heterogeneous forecast practices by customers and diversified demand portfolio.

$Q_{ft-k,t}$ = demand forecast quantity reported in period $t-k$ for demand in period t, i.e., this is the lead-time k forecast. We observe a pattern in the forecast data where for a random CUST-P/N pair and for any given value of t, the values of $Q_{ft-k,t}$ tend to increase as k increases from $k=-12$ to $k=-6$ and then gradually decrease again. Also, note that the standard deviation of CUST-P/N forecast in each period is significantly higher compared to the average forecast a random customer-P/N might make, suggesting the presence of a wide range in the magnitude of CUST-P/N forecasts. The final forecast, $Q_{ft,t}$ is denoted also as the customer request (Q_r). The supplier response to this customer request Q_r is the supplier's commit quantity (Q_c).

5. Data Transformation and Variable Description

In this section we describe the creation of new variables from the raw data in Table 1. In subsequent sections, we suppress the subscripts i and j (which denote customer and P/N, respectively) unless the context requires such specification. We adopt the same terminology used by Heath and Jackson (1994). To facilitate statistical analysis, we aggregated all monthly data into quarterly data after performing extensive data quality procedures.

5.1. *Forecast Series and vectors:*

At the beginning of each time period, customers forecast their anticipated demand for the upcoming H periods, We represent these forecasts using a "forecast vector" at time t of length H: $\{Q_{ft,t}, Q_{ft,t+1}, \ldots, Q_{ft,t+H}\}^{T}$ where, as described in Section 4, $Q_{f\,s,t}$ denotes a demand forecast made in period s for anticipated demand in period $t \geq s$ (and $\{\ \}^T$ denotes the transpose of the vector).

Consider the set of four forecast vectors with length $H=5$, depicted in Figure 3a. The first vector is populated at the beginning of period 1, the second is populated at the beginning of period 2, …, and the fifth at the beginning of period 5. Since the second forecast vector is populated at the beginning of period 2, its first element is $Q_{f_{2,2}}$, namely, the forecast made at the beginning of period 2 for delivery in period 2. Thus, $Q_{f_{2,2}} =$ $Q_r(2)$, i.e., the forecast made at t = 2 for a delivery in 2 is the actual request made at t = 2. For the purpose of recognizing requests as final forecasts (e.g., following H-1 forecasts in the preceding *H-1* periods), the horizon is technically considered as an *H*-period horizon (although it only covers *H-1* periods of calendar time). In our analysis, we will be concerned both with horizons of $H = 5$ and $H = 13$, corresponding to dividing the 12-month period preceding final orders into quarterly or monthly periods. Notice that as one moves up and to the right along the minor diagonals in the matrix in Figure 3, the second subscript remains constant and the first subscript increases. This represents forecasts made in increasing periods for a demand in a given future period. We refer to this sequence of *H* forecasts $\{Q_{ft-H+1,t}, Q_{ft-H+2,t}, \ldots, Q_{ft,t}\}$ made in increasing periods for demand in a given time period *t* as a "forecast series".

5.2. Churn:

Churn is the change in the forecast values between any two forecasts for a given final period demand in a forecast series. Formally, let $\{Q_{ft-H+1,t}, Q_{ft-H+2,t}, \ldots, Q_{ft,t}\}$ be a forecast series of length H for an order to be delivered in period t; then churn is defined, for $s < t$ and $s = t - H+1, \ldots, t-1$, as:

$$C_{s,t} = \frac{Q_{f_{s+1,t}} - Q_{f_{s,t}}}{Q_{f_{s,t}}}. \qquad (1)$$

Churn, as defined above, represents the percentage change in forecasts. It is synonymous with the multiplicative model of forecast evolution used by Heath and Jackson (1994). Note from the above definition that churn can be either positive or negative.

In parallel to our definitions of forecast series, the corresponding churn series is defined as the sequence of churn quantities along the minor diagonal and a churn vector as the corresponding vertical sequence. For example, in Figure 3a forecast vector of length five at time $t=2$ is the second column containing the elements $\{Q_{f22}, Q_{f23}, Q_{f24}, Q_{f25}, Q_{f26}\}^T$. Similarly, the forecast series of length five at $t = 1$ is $\{Q_{f15}, Q_{f25}, Q_{f35}, Q_{f45}, Q_{f55}\}$. For forecast series and vectors of length *H*, churn series and vectors will be of length *H-1*.

5.2.1. Positive and Negative Churn:

Measurement of churn is important to the supplier, as it is this variability and uncertainty in customer forecasts that is the greatest contributor to the supplier's risk. We decompose churn into its positive and negative components, defining $C^{+}_{s,t} = Max[C_{c,t}, 0], C^{-}_{s,t} = Min[C_{s,t}, 0]$. (See Figure 3c.) Positive churn represents the month-over-month increase in forecasted demand for a given month *t* and a negative churn captures the month-over-month decrease in forecasted demand for a given month *t*. We perform separate analysis on these two portions of churn because analysis of the economics of the semiconductor industry as well as discussions with management at the manufacturing facility revealed that positive and negative churn can have different impacts on the supply chain. As an example, the 12-component churn vector $C = \{0, 0.1, 1, -0.55, 0.2, -0.2, -0.2, 0.3, 0.4, -0.5, -0.1, -0.1\}$ is decomposed into positive churn (C^{+}) and negative churn (C^{-}) as follows: $C^{+} = \{0, 0.1, 1, 0, 0.2, 0, 0, 0.3, 0.4, 0, 0, 0\}$ and $C^{-} = \{0, 0, 0, -0.55, 0, -0.2, -0.2, 0, 0, -0.5, -0.1, -0.1\}$. Notice that $C = C^{+} + C^{-}$.

5.2.2. Quarterly churn:

In our dataset, the generic structure of a typical data element had twelve monthly forecasts leading to an order, i.e., thirteen entries in all. This gave rise to a forecast vector of length thirteen which gave rise to twelve month churn vectors. We note that some of the forecast data was missing; when it is, we simply assume that the missing forecast is equal to the last forecast made prior to the missing forecast. If the first forecast made occurs after $t - 12$ (i.e., START for a given CUST-P/N pair is greater than -12), then the missing forecasts from $t - 12$ until $t - START$ are set equal to 0. However, sales and operations planning departments in the semiconductor industry often plan based on *quarterly* aggregate forecasts and churn quantities. Therefore, we aggregated all data to the quarterly level. For instance, the positive vector $C^{+} = \{0, 0.1, 1, 0, 0.2, 0, 0, 0.3, 0.4, 0, 0, 0\}$ was converted to $\{1.1, 0.2, 0.7, 0\}$ and the negative vector $C^{-} = \{0, 0, 0, -0.55, 0, -0.2, -0.2, 0, 0, -0.5, -0.1, -0.1\}$ was transformed to $\{0, -.75, -0.2, -0.7\}$.

From here on, since we deal only with quarterly churn series and vectors, we use a terminology that is easier to interpret. We obtain quarterly churn by first computing monthly churn for the 12-month period preceding a given order. This leads, per the above discussion, to churn vectors and churn series, each of length 12. From these we compute from these the positive and negative monthly churn vectors, also of length 12. We then aggregate these positive and negative churn vectors and series into 4 quarterly (positive and negative) churn vectors and series by summing the respective monthly periods in each quarter. This leads to the following definitions:

$Cv^{-}q_k(t)$ = Aggregate negative churn at month t for quarter $k = 1, 2, 3, 4$ in the future

$Cv^{+}q_k(t)$ = Aggregate positive churn at month t for quarter $k = 1, 2, 3, 4$ in the future

where

$$Cv^{-}q_k(t) = C^{-}_{t,t+3(k-1)+1} + C^{-}_{t,\ t+3(k-1)+2} + C^{-}_{t,\ t+3(k-1)+3}\ ,\ k = 1, 2, 3, 4$$

$$Cv^{+}q_k(t) = C^{+}_{t,\ t+3(k-1)+1} + C^{+}_{t,\ t+3(k-1)+2} + C^{+}_{t,\ t+3(k-1)+3}\ ,\ k = 1, 2, 3, 4$$

Similarly, quarterly churn *series* are defined as quarterly aggregates of positive and negative churns for month t, as made k quarters in advance of month t. More specifically, we define:

$Cs^{-}q_k(t)$ = Aggregate negative churn for month t as determined by forecasts made in quarter $5\text{-}k$, $k = 1, 2, 3, 4$ preceding t.

$Cs^{+}q_k(t)$ = Aggregate positive churn for month t as determined by forecasts made in quarter $5\text{-}k$, $k = 1, 2, 3, 4$ preceding t.

where

$$Cs^{-}q_k(t) = C^{-}_{t-3(5-k),t} + C^{-}_{t-3(5-k)+1,t} + C^{-}_{t-3(5-k)+2,t}\ ,\ k =1, 2, 3, 4$$

$$Cs^{+}q_k(t) = C^{+}_{t-3(5-k),t} + C^{+}_{t-3(5-k)+1,t} + C^{+}_{t-3(5-k)+2,t}\ ,\ k =1, 2, 3, 4$$

Thus, $Cv^{-}q_1(t)$, $Cs^{+}q_3(t)$ denote aggregate negative churn *at time t* in the 1st quarter and aggregate positive churn *for time t* in the 3rd quarter, respectively. $Cv^{-}q_1(t)$ is the churn at time t for $\{t+1, t+2, t+3\}$ and $Cs^{+}q_3(t)$ is the churn for time t at $\{t-6, t-5, t-4\}$. Moreover, note that the churn quantities are aggregate quantities for the quarter, but the index t is denominated in months (namely, the monthly order quantity to which the underlying quarterly aggregates are targeted). To emphasize the difference between series and vectors, note that a churn series is a forecast update *for time t* that occurs along the *minor* *diagonal* of a churn matrix, (see Figure 3) while a churn vector component is a forecast update that occurs *at time t* along the *vertical column* of a churn matrix.

To summarize, the raw forecast data denoting *13 element rolling monthly forecast* vectors (of which the final component is the actual order) were transformed into corresponding 12-month churn factors, which were transformed into their positive and negative parts. These latter were further transformed to *positive and negative 4-element quarterly churn* vectors by aggregating positive and negative churn within each of the respective quarters.

5.3. Supply Allocation and Rationing:

Customer i requests quantity Q_{rt} of P/N j at time t to be delivered on a specific delivery date. The supplier responds with a commit quantity Q_{ct}. It is not always the case that $Q_{rt} = Q_{ct}$. For example, if the supplier is facing constrained capacity he may be unable to meet all customer demand, in which case the supplier's commit quantity may be less than the customer's request quantity.

We now define two related variables: allocation (*alloc*) and supplier allocation (*Falloc*). Allocation is defined at the customer level. It represents the fraction of the customer's request that the supplier commits to deliver. $Falloc_t$ is defined at the aggregate supplier level. It represents the fraction of *all* orders for time period t that receive an allocation strictly less than one.

More specifically, for customer i and P/N j, allocation for a request made at time t is the fraction of the customer's request that the supplier commits to deliver. Allocation is given by:

$$alloc_{ijt} = \frac{Q_{cijt}}{Q_{rijt}}.$$

If the supplier commits to deliver to the customer everything that the customer requests, then the customer receives a full allocation and $alloc_{ijt} = 1$. If the supplier is unable to meet the customer's full request quantity, then allocation will be less than 1. In some cases it is possible that the allocation will exceed 1. For example, the supplier may wish to smooth production and will encourage the customer to receive product earlier than the customer had planned and requested.

Supplier level allocation in period t is defined as the fraction of all orders in period t that do not receive a full allocation.

$$Falloc_t = \frac{\left\| \left\{ (i,j) \mid alloc_{ijt} < 1 \right\} \right\|}{\left\| \left\{ (i,j) \mid Q_{rijt} \neq 0 \right\} \right\|}.$$

6. Episodic Forecasting Behavior and Hypotheses Development

Forecasting Behavior

In this section, we describe a phenomenological observation about the forecasting behavior of the customer pool as observed by the supplier's manufacturing facility. Figure 4a shows the time series plot of the quarterly churn series; $Cs^{-}q_1(t)$, $Cs^{-}q_2(t)$, $Cs^{-}q_3(t)$ and $Cs^{-}q_4(t)$. We observe that the four peaks in Figure 4a are equally spaced by 3 months (in other words a "quarter"). Hence, if we displaced the 1st quarter churn by 9 months, the second quarter churn by 6 months and the 3rd quarter churn by 3 months, all the peaks graphically align themselves in a straight vertical line. Not surprisingly, the lagging process as described above transforms the churn series to churn vectors. To avoid visual misperceptions, we show the corresponding correlation matrix (i.e., quarterly churn vector components) in Table 2b. Table 2a shows the correlation matrix of the four quarterly churn series components. Comparing the correlations between the churn vectors (Table 2b) with the corresponding entries in the churn series (Table 2a), we see that the entries in Table 2b are significantly higher. For instance, the correlation between $Cs^{-}q_4(t)$ and $Cs^{-}q_2(t)$ is -0.10 while the correlation between $Cv^{-}q_4(t)$ and $Cv^{-}q_2(t)$ is 0.65.

Figure 4b and Table 2b taken together suggest that the four components of the quarterly churn vector are modified almost equally, i.e., the information incorporated in each of the 12 monthly forecast updates is identical. This behavior, although non-intuitive at first, is consistent with the predictions of the MMFE model (see Heath and Jackson, 1994). The high correlation between the components of the churn vector ***at*** time t, both for positive and negative churn, suggests that customers observe episodes of supply and demand shocks that affect their perception of ***all future capacity requirements***. We refer to this forecast updating process as "episodic forecast updates".

For the remainder of the paper, we investigate the underlying mechanics of the supply and demand process that produce this correlation matrix (in Table 2b). In practice, speaking to managers, forecast updates are not necessarily reflective of incremental information about customers' demand. The forecasts are, in fact, an outcome of a decision model used by customers; a decision model that seeks to minimize the cost of supply and demand mismatch for a given finite horizon (12 months, in most cases).

We take a closer look at the notion of perceived mismatch between supply and demand in rolling horizons (12 months to be precise); a multi-period setting, as causing the episodic behavior. The following section describes in detail the development of hypotheses deriving from the above observations.

Hypotheses Development

Customers receive common signals about demand conditions, supply conditions, and macroeconomic forecasts. They also receive a private signal via the supplier's allocation policy. In our data, the supplier did not disclose his allocation policy, but only revealed the commit quantities once customer requests were made. We therefore use the allocation variables ($alloc_{ijt}$) as a proxy for the customer's private supply updates. The unintended consequences of the allocation policy are reflected broadly in the customer orders and forecast churn. In the spirit of Schelling (1978), we hypothesize about both the macrobehavior, i.e., orders at the level of the supplier's manufacturing facility (churn), and the possible distribution of micromotives, i.e., the miasma of CUST-P/N level behaviors, that could in aggregate lead to the macrobehavior we observe in the data (Figure 4a, Figure 4b).

Orders: We conjectured initially that, at the supplier level, all unfulfilled requests from customers would reappear as inflated requests in future periods. Thus, we expected customers would be backlogged when they were allocated less than 100% of their order. However, the data indicates that some unfilled customer requests are backlogged while others disappear from the system altogether. The backlogging may be performed by the customer or by the supplier when he sees that supply is constrained. Alternative explanations for unfilled demand that disappears from the system could be that some customers leave the supplier and go to a competitor or that the supplier may decide to drive part numbers out of the system because they are too costly to produce relative to their contribution to the bottom line.

At the micro level, each CUST-P/N combination is treated differently by the supplier. Customers may be assigned different priority in allocation, due to such factors as the size of the customer, the nature of the customer's relationship with the supplier, or due to other considerations. Customers typically order multiple P/Ns from the supplier. We anticipated that the supplier sets allocation priorities based on relationships with customers and that the individual part numbers, to the extent that the part numbers face similar capacity constraints, do not affect this allocation priority. (Clearly, if specific part numbers face particular capacity constraints for particular resources, then some (unconstrained) part numbers will not be affected while others will. In short, the part number may play a role in the supplier allocation priority decisions.)

Churn: Following our observation of the episodic behavior (in section 4), it follows that all components of a churn vector would respond to episodes of allocation. In the remaining part of this section, churn refers to the entire quarterly churn vector (four components), unless otherwise specified.

-Positive Churn: While backlogs may manifest themselves as increased orders in future periods, customers may also over-react to rationing and inflate future orders well beyond the direct backlogged orders (see also Watson and Zheng, (2005)). Given our empirical observation of episodic behavior, we conjecture that churn would increase at the supplier level following a period of heavy rationing. Even though customers' underlying forecasts may not be correlated, their forecast updates may become correlated because they share a *common rationing episode*. Industry analysts commented to the authors[2] that episodes of

---

[2] Thus, Semico Research Corporation uses several methods of forecasting tight capacity. Some of these are based directly on major customer demand forecasts for specific semiconductor demand segments.

capacity allocation are often signaled by a group of high priority customers increasing their forecasts. A possible scenario is that the supplier reacts by allocating a larger share of the capacity to this high priority group and consequently allocates a smaller share of the production capacity to the lower priority customers. When faced with this allocation episode, the low priority customers could then over-react and increase all their forecasts in the belief that they would get their cumulative demand satisfied at some point within the next 12 months. Building on this scenario where a common rationing episode is initiated by demand spikes of the high priority customers but the shocks spill over to the low priority customers' forecasts, we refer to such spillover effects as *contagion*. Contagion would manifest itself as a correlation between seemingly uncorrelated "demand" forecasts driven by episodes of capacity allocation.[3] Contagion can be detected by observing correlations between forecast updates (positive churn) ex-post of an allocation episode. An unfortunate consequence of contagion, when forecast updates reflect positive correlation across CUST-P/N combinations, is that the supplier is unable to benefit from *risk pooling* of customer orders. We will examine the existence and magnitude of contagion at the CUST-P/N (micro) level, with particular attention paid to whether the contagion is evident at the P/N level, the customer level, or both.

-Negative Churn: At the level of the supplier's manufacturing facility, one anticipates temporal shifts in forecasts, i.e., production smoothing following an allocation episode. Production smoothing is sometimes initiated by the supplier in coordination with customers. Discussions with industry managers suggested that end-of-quarter sales loading effects were very much in evidence in the industry. These effects, if present, would lead to negative, respectively positive, churn at the beginning, respectively end, of a quarter in meeting or exceeding financial targets. The supplier studied did use quarterly targets during the study period, to track financial and sales performance. Hence, we expected to see these end-of-quarter effects. Negative churn could also increase when customers leave the supplier's manufacturing facility, with all their existing forecasts then terminated.

---

[3] The phenomenon of forecast inflation in anticipation of allocation in a single-period setting is discussed by Cachon and Lariviere (2001).

7. Econometric Models

In this section, we lay out the econometric models to test our hypotheses about customer orders and churn and contagion. As outlined in the previous section, episodic behavior manifests itself in the form of customers' orders, positive churn and negative churn; at both at the level of the supplier's manufacturing facility (macrobehavior) and at the CUST-P/N level (microscopic).

7.1. Macrobehavior Models

We begin our analysis by considering the impacts of allocation policies on aggregate forecasts at the supplier's manufacturing facility level. The observed churn effects, if any, associated with allocation may be thought of as the macrobehavioral outcomes of individual customer decisions as these interact with the supplier's capacity and demand management processes.

Macrobehavior Orders:

$$\left(\sum_{i,j} Q_{rijt}\right) \sim M1 + M2 + \alpha_0 \left(\sum_{i,j} Q_{rijt-1}\right) + \alpha_1 Falloc_{t-1} + \varepsilon_t \qquad (2)$$

The LHS of (2) represents total requests to the supplier's manufacturing facility at time *t*. We expect it to be affected positively by the previous period's aggregate requests and show a positive reaction to the rationing at *t-1*; if more customers are rationed then aggregate customer requests in the next period is larger. *M1* and *M2* are monthly dummies to capture the beginning and end-of-quarter effects. We estimate the above equation using a simple LSDV (least squares dummy variable) estimator.

Macrobehavior of positive churn

The positive churn vectors exhibited episodic behavior. (See Figure 3b, Table 2b.) Here, we quantify the effect of this positive churn at the supplier manufacturing facility level, by correlating against its first-order autoregressive (*AR(1)*) component and the strength of rationing at *t-1*.

$$\left(\sum_{i,j} Cv^{+} q_{ijkt}\right) \sim M1 + M2 + \chi_{0k}\left(\sum_{i,j} Cv^{+} q_{ijkt-1}\right) + \chi_{1k} Falloc_{t-1} + \varepsilon_t,\ k = 1,2,3,4 \quad (3)$$

Consistent with the idea that all the churn components are affected by the common rationing variable, we estimate equations (2) and (3) independently of each other. Positive and statistically significant coefficients on χ_{11}, χ_{12}, χ_{13} and χ_{14} would confirm our contagion hypothesis, i.e., rationing episodes cause a large scale cross sectional forecast distortion through positive churn; a sign that forecast churn across customers would increase in tandem following the rationing episode. *M1* and *M2* are monthly dummies to capture the beginning and end of quarter effects. We use a LSDV estimator to estimate each of the equations.

Macrobehavior of negative churn

Similar to the positive churn formulation, we use four independent linear equations to model the macro behavior at the level of the supplier's manufacturing facility vis-à-vis negative churn. However, we also include an additional equation that models the process of customers exiting the facility following periods of rationing.

$$\left(\sum_{i,j} Cv^{-} q_{ijkt}\right) \sim M1 + M2 + \eta_{0k}\left(\sum_{i,j} Cv^{-} q_{ijkt-1}\right) + \eta_{1k} Falloc_{t-1} + \varepsilon_t, k = 1,2,3,4 \quad (4)$$

$$\left(\sum_{i,j} sysexit_{ijt}\right) \sim M1 + M2 + \psi_0\left(\sum_{i,j} sysexit_{ijt-1}\right) + \psi_1 Falloc_{t-1} + \varepsilon_t \quad (5)$$

We expect the coefficients of $Falloc_{t-1}$ in equation (4) to be negative, i.e., as the intensity of supply allocation increases, negative churn gets worse. However, we expect ψ_1 in equation (5) to be positive, i.e., as supply allocation gets more intense, the number of CUST-P/N exiting the supplier's manufacturing facility would also increase. *M1* and *M2* are monthly dummies to capture the beginning and end of quarter effects. We use a simple LSDV estimator to estimate (4) and (5).

7.2. Micromotive Models

In this section, we consider the impact of rationing on individual CUST-P/N forecast churn. This reflects the consequences of rationing policies on individual orders.

Micromotives of customers' orders:

The customers and the supplier interact repeatedly in each period. At the beginning of each period *t* customers place their requests, with the knowledge of the allocation they received in time *t-1*. The supplier sees the request quantities and responds with a commit quantity Q_c.

The customer's reaction to allocation is modeled as:

$$Qr_{ijt} \sim \alpha_0 Qr_{ijt-1} + \left(\tilde{\alpha}_{1ij}\right) alloc_{ijt-1} + \varepsilon_{ijt} \tag{6}$$

The supplier's reaction to a customer request is modeled as:

$$alloc_{ijt} \sim \beta_0 alloc_{ijt-1} + \left(\tilde{\beta}_{1ij}\right) Qr_{ijt} + \delta_{ijt} \tag{7}$$

In (6), $\tilde{\alpha}_{1ij}$ is the regression coefficient in the customer's response equation that describes the sensitivity of the CUST-P/N combination to allocation in time *t-1*. We model $\tilde{\alpha}_{1i,j}$ as a random effect to capture the notion that CUST-P/N response to allocation could be drawn from a probability distribution. (See Maddala, 1992 for details.) Without any additional information on the natural sampling process in operation, we desist from using other alternatives such as a fixed effects model or a hierarchical linear model. Similar to $\tilde{\alpha}_{1i,j}$, we also model $\tilde{\beta}_{1ij}$ as a random effect. $\tilde{\beta}_{1ij}$ models the supplier's allocation policy as a random draw from a probability distribution. Note that both $\tilde{\alpha}_{1ij}$ and $\tilde{\beta}_{1ij}$ are random effects that could also be modeled as $\tilde{\alpha}_{1i}$ and $\tilde{\beta}_{1i}$ (independent of P/N) or $\tilde{\alpha}_{1j}$ and $\tilde{\beta}_{1j}$ (independent of CUST). We indeed estimate the models using these alternative random effects to assess whether CUST, P/N or CUST-P/N best explain cross-sectional variation.

Both the customers' and supplier's responses require simultaneous estimation. Therefore, we use a vector auto-regression (VAR) model. (See Enders, 1995 for related theory.) The two equations represent a structural VAR. Since, we cannot estimate this structural VAR as a Seemingly Unrelated Regression (SUR) Model, by using a transformation, we convert the equations to a reduced form VAR (RVAR) as follows.

$$Qr_{ijt} \sim \alpha_0 Qr_{ijt-1} + \left(\tilde{\alpha}_{1i,j}\right) alloc_{it-1} + \varepsilon_t \tag{8}$$

$$alloc_{ijt} \sim \left(\alpha_0 \tilde{\beta}_{1ij}\right) Qr_{ijt-1} + \left(\beta_0 + \tilde{\alpha}_{1ij}\tilde{\beta}_{1ij}\right) alloc_{it-1} + \left(\delta_t + \varepsilon_t \tilde{\beta}_{1ij}\right) \tag{9}$$

The decoupled equations are treated as seemingly unrelated in the reduced form VAR and can now be estimated independently of each other. The customer's response, given by (8), is estimated using an appropriate random effects MLE estimator. The supplier's reduced form (9) is also estimated using a random effects estimator.

Micromotives of positive and negative forecast churn

$$Cv^{+}q_{ijkt} = \lambda_{0k} Cv^{+}q_{ijkt-1} + \tilde{\lambda}_{1ijk} alloc_{ijt-1} + \varepsilon_{ijkt}, \quad \forall\text{(i,j)};\ k = 1,2,3,4 \tag{10}$$

$$Cv^{-}q_{ijkt} = \varpi_{0k} Cv^{-}q_{ijkt-1} + \tilde{\varpi}_{1ijk} alloc_{ijt-1} + \varepsilon_{ijkt}, \quad \forall\text{(i,j)};\ k = 1,2,3,4 \tag{11}$$

As in the micromotives model of order backlogging (6), we use random effects to capture the reaction of customer churn to allocation. Churn (both positive and negative) is expected to have a *AR(1)* component and as before, the random effects (coefficient of $alloc_{i,j,t-1}$ depend on *i* (CUST) and *j* (P/N). Similar to the order batching model, we could model the supplier's reaction to churn as well. However, as we saw in the order backlogging micromotives model, the reduced form buyer equation (8) was identical to the SVAR equation (6). Since our interest centers on the customer's response function, for the sake of brevity[4], we do not explain the VAR form of the model.

8. Results

The results from the estimation of macro behavior models ((2)-(5)) are summarized in Table 3 and the results from the micromotives models ((6)-(11)) are summarized in Tables 4a, 4b and 4c.

[4] In a forecasting model of supplier and customer reactions to allocation and churn respectively, we would use the full fledged VAR model. For illustrating micromotives, the single equation model suffices.

Macrobehavior of Request Quantities Q_r

The request quantities/orders are modeled using (2) and the results are reported in Table 3. The presence of dummy variables necessitates the use of a LSDV for this model. The *AR(1)*, $Falloc_{t-1}$ and the dummy for beginning quarter (*M1*) are statistically significant at the 1% level and the second dummy (*M2*) is significant at the 5% level.

The coefficient of $Falloc_{t-1,}$ i.e., α_1, is positive and significant at the 1% level. Since $Falloc_{t-1}$ measures the percentage of customers who receive less than 100% allocation, this result is consistent with the notion of customers carrying backlogs from a previous period into the next period. The correlation between Q_{rt} and $Falloc_t$ is 0.09, which is consistent with the hypothesis that the facility does not respond to increased requests in one period by increasing capacity in the following period.

The *AR(1)* coefficient i.e. α_0, is 0.8037, suggesting behavior close to unit root. The mean reversion exhibited by the request quantities is quite mild compared to the churn time series which exhibit smaller *AR(1)* coefficients (discussed in more detail in the next section).

The monthly dummies are significant at the 1% and 5% levels, respectively. The negative coefficients suggest that the requests consistently drop in the beginning and the middle of a quarter and consistently spike upwards at the end of each quarter. One would therefore expect to a see a spike in orders in March, June, September and December. This could be explained by the fact that the supplier operates financially on a calendar quarter basis. Thus, the data may reflect an incentive by the supplier to displace demand forward at the end of the quarter (e.g., by offering discounts) in order to meet projected numbers.

Finally, the high R^2 (95.02%) suggests excellent fit. The residuals from the regression were passed for normality using the Kolmogorov-Smirnoff test and the Chi-square tests. The Durbin-Watson statistic (1.96) for the data suggests negligible auto-correlation in residuals. Estimating the model as a simple ordinary least squares (OLS) without the dummy variables reduced the R^2 to 75.6%, but the residuals showed partial auto-correlations at a lag of 3 months (0.67) and 6 months (0.45).

Macrobehavior of Positive Forecast Churn ($Cv^+q1(t)$, $Cv^+q2(t)$, $Cv^+q3(t)$, $Cv^+q4(t)$)

The churn statistics modeled using (3) were estimated using LSDV. The *AR(1)* coefficients are significant at the 5% level for the first 3 quarters, but the 4th quarter *AR(1)* term is not statistically significant. Coefficient of $Falloc_{t-1}$ is significant for all quarters at the 5% level. The statistical significance of the monthly dummies (*M1*, *M2*) does not show any consistent patterns.

The coefficient of $Falloc_{t-1}$ increases from first quarter (0.0431) to fourth quarter (0.3897). Figure 2a and Figure 2b depict this effect for negative churn. The consistent effect of $Falloc_{t-1}$ on every quarter of positive churn, suggests that customers are over-reacting to supply allocation. One might expect customers to increase forecasts only in the short run, leaving the long run forecasts to adjust to macroeconomic factors. However, the impact of $Falloc_{t-1}$ increases with each quarter, so that $Cv^+q4(t)$ is more sensitive to $Falloc_{t-1}$ than $Cv^+q1(t)$.

The *AR(1)* coefficients for positive churn suggest that the further out a month is the more volatile its forecasts would be. This behavior can be explained by Figure 1 which lays out the timeline of customer-supplier interaction. Note that at time *(t-12)* customers are speculating on the entire order they place at the facility which is whetted by the possibility that they might even cancel the order. Closer to the order date *(t)*, the short run forecasts are likely to be focused on fine tuning existing current forecasts.

The monthly dummies once again show consistently positive sign while their statistical significance does not exhibit any visible pattern across quarters. The positive sign suggests that the beginning of every quarter is characterized by an increase in positive churn.

Finally, the high R^2 on all the models suggests excellent fit. The residuals from the regression were passed for normality using the Kolmogorov-Smirnoff test and the Chi-square tests as well for normality. The Durbin-Watson statistic for the residuals suggests negligible auto-correlation in residuals. When the dummy variables were removed and the model was estimated using simple OLS, the partial autocorrelation function across the residuals suggested strong quarterly seasonal effects.

Negative Churn (*$Cv^{-}q1(t)$, $Cv^{-}q2(t)$, $Cv^{-}q3(t)$, $Cv^{-}q4(t)$*)

We model the churn statistics using model (4). The presence of dummy variables necessitates the use of LSDV for this model. The *AR(1)* coefficients are significant at 1% level for all four quarters. The coefficient of $Falloc_{t-1}$ is significant for all quarters at the 5% level. The monthly dummy for the beginning of a quarter (*M1*) is significant at the 5% level in all quarters while the second dummy (*M2*) is only significant at the 10% level for $Cv^{-}q1$.

The effect of $Falloc_{t-1}$ on negative churn is quite pronounced. The magnitude of negative churn increases (negative regression coefficients) when the number of customers rationed increases. This effect is more pronounced in the medium and long run ($Cv^{-}q3$, $Cv^{-}q4$) compared to the short run forecasts ($Cv^{-}q1$, $Cv^{-}q2$). While the macrobehavior captures the aggregate effect, the underlying micromotives of customers are not revealed by this result. Unlike positive churn which reflects only forecast increases, negative churn is influenced by forecast changes, sysexits, and batching of forecasts.

The *AR(1)* coefficients are all positive and significant at the 1% level. Mean reversion is strongest in the first quarter (short run forecasts) and weakest in the third quarter (medium run). Negative churn differs from positive churn in what it measures, i.e., negative churn can be produced by customers exiting the system and/or batching their existing forecasts.

The monthly dummy for beginning of quarter (*M1*) is significant at the 1% level in all quarters. The negative sign on *M1* in all quarters suggests that the magnitude of negative churn increases at the beginning of each quarter. Dummy variable *M2* is consistently smaller by a factor of approximately 10, but statistically insignificant nonetheless.

SystemExits (sysexit)

In Table 1, we observed that roughly 70% of the cross section leaves the supplier's manufacturing facility. Thus, it is not clear if supply allocation is driving this behavior. Since negative churn also measures customers exiting the system, we test for the effect of allocations on sysexits. We find that sysexits shows a strong mean reversion (*AR(1)* = 0.6399) and a statistically significant correlation with $Falloc_{t-1}$, suggesting that periods of high supply allocation are followed by periods of increased CUST-P/N exiting the facility. The monthly dummies (*M1*) are also significant at the 1% level, which is consistent with managers' opinions that the supplier's decisions may be influenced by quarterly financial and operational targets.

Micromotives

The estimation of the micromotives models is reported in Tables 4a, 4b and 4c. As discussed earlier, we report the VAR formulation only for the request quantities in Table 4a.

Request Quantities (Q_r)

We estimate the stochastic coefficients (mixed linear model) embedded in a VAR formulation. Since the stochastic coefficients depend on the cross-section, the cross section is specified in three different ways- CUST, P/N and CUST-P/N pair. The results of the estimation from the three different methods are described in Table 4a. In addition to the cross sections, Table 4a also specifies the 2-equation VAR estimation for request quantity (Q_r) and *alloc*.

The ANOVA comparison of the three cross sectional models (6)-(11) suggests that the CUST-P/N model is more appropriate for analysis as it better explains between-group variance. For Q_r, the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) for the RVAR estimator is lowest for the CUST-P/N model. The Log likelihood is also greater for the CUST-P/N model compared to the CUST and P/N specifications. For further analysis, we continue with the interpretation of Model (6) (specified by CUST-P/N).

Customer's actions (Q_r): The estimation results for the RVAR equation hold for the SVAR model as well. The fixed effect response to allocation is positive (0.0214), i.e., the data is consistent with customer orders decreasing in response to allocation. This appears to contradict the finding that unfilled customer orders are backlogged at the microlevel and is also in contrast to the supplier's manufacturing facility level macro model. (See Table 4a, coefficient of $alloc_{t-1}$.) This apparent contradiction is explained by the random effects coefficient of $alloc_{t-1}$=0.1901. This suggests that customer's response to allocation in period *t-1*; $\tilde{\alpha}_{1i,j} \sim$ N(0.1024, 0.1901^2). While some customer orders are backlogged in response to allocation (a negative draw from this distribution), other customer orders are reduced or cancelled in response to allocation (positive draws from this distribution). The probability that a customer order is backlogged in response to allocation is calculated as $\Pr\left(\tilde{\alpha}_{1i,j} < 0\right)$=0.46. Thus, a significant fraction of unfilled customer requests (46% of the CUST-P/N population) is backlogged into the following period. Both the fixed effect terms are significant at the 1% level. The *AR(1)* coefficient (0.1024) suggests strong mean reversion properties at the microlevel. This is in contrast to the mild mean reversion property at the level of the supplier's manufacturing facility (*AR(1)* coefficient was 0.8037).

Supplier's actions (*alloc*): Analyzing the supplier's response to customers' requests is a little more involved because the stochastic effects are embedded on three independent variables. The fixed effect terms in the stochastic RVAR are all significant at the 1% level. The supplier's response to customer requests depends on the CUST-P/N and this stochastic nature of $\tilde{\beta}_{1i,j}$ is estimated by the RVAR as N(0.1843, 0.273^2), i.e., $\tilde{\beta}_{1i,j}$~N(1.798, 2.66^2). The normal distribution assumption notwithstanding, the probability that the supplier responds to increased requests by rationing the CUST-P/N is calculated as $\Pr\left(\tilde{\beta}_{1i,j} < 0\right)$=0.26, i.e., roughly 26% of the population gets rationed (lower priority).

The intercept is N(-0.0051, 0.004^2), suggesting that the intercept is almost a constant and the stochastic coefficient is unnecessary. The *AR(1)* coefficient is (-0.046, 0.202^2) which in the SVAR form translates to an expected *AR(1)* coefficient of -0.0043. The mild negative sign on the *AR(1)* coefficient agrees with our intuition of allocation having a strong mean reversion (almost oscillatory behavior around the mean).

To summarize, ex-post linear regression analysis reveals that 26% of the CUST-P/N population in the data gets rationed, while roughly 46% of the population backlogs in response to order rationing. Ex ante, the supplier is not aware which 46% of the population is prone to backlogging. Similarly, the customers (CUST-P/N to be precise) are also not privy to their position in the allocation policy either.

Positive Churn (*$Cv^{+}q1(t)$, $Cv^{+}q2(t)$, $Cv^{+}q3(t)$, $Cv^{+}q4(t)$*)

For each of the four quarters, we estimate the stochastic coefficient models using a restricted maximum likelihood estimator. The results of the estimation are summarized in Table 4b. We consider three forms of stochastic settings- CUST, P/N and CUST-P/N, as a basis for the stochastic coefficients. We compare the models based on their AIC, BIC, and Log Likelihood (see Maddala, 1992 for a detailed discussion of AIC, BIC and Log Likelihood goodness of fit measures.) We find that the model with random effects specified by P/N has the best fit.

The stochastic coefficients in quarters $k = 1, \ldots, 4$ are estimated to be N(-0.0129, 0.1808^2), N(-0.00054, 0.0007^2), N(-0.02, 0.0553^2) and N(-0.0091, 0.0316^2), respectively. The random coefficients suggests that roughly 53%, 78%, 63% and 59% of the customers react to allocation in period *t-1* by increasing their churn in quarters $k = 1, \ldots, 4$, respectively. The almost uniform increase in every quarter's churn points to a propensity among customers to overreact to episodes of supply-demand mismatch (*alloc*) by increasing their positive churn, i.e., by increasing forecasts. The exaggerated positive shifts in every quarter's churn in response to allocation are also observed at the supplier's manufacturing facility level, thus confirming our hypothesis of micro-level over-reaction and macro-level contagion.

The *AR(1)* components in each quarter are significantly smaller in magnitude compared to the macrobehavior model. Autoregression is weakest in the fourth quarter (0.1154) and strongest in the second quarter (0.3546), mirroring the macrobehavior predictions. Thus both the micromotives and the macrobehavior models suggest that the uncertainty in long term forecasts is greater than the short and medium term forecasts.

Negative Churn (*$Cv^{-}q1(t)$, $Cv^{-}q2(t)$, $Cv^{-}q3(t)$, $Cv^{-}q4(t)$*)

Similar to the positive churn models, we estimate three models, specified by CUST, P/N, and CUST-P/N and choose the model with best fit based on AIC, BIC and LogLikelihood values for these models. We find that a random effect specified by P/N has the best fit for the model.

Negative churn, unlike positive churn, is a manifestation of more than simple forecast decreases. Negative churn can also be produced by sysexits and forecast batching. The stochastic coefficients for $alloc_{t-1}$ are estimated to be N(0.0212, 0.2402^2), N(0.0136, 0.1774^2), N(-0.0062, 0.0081^2), N(0.0749, 1.011^2) which translates to roughly 54%, 53%, 22% and 52% of the customers increasing the magnitude of their negative churn in response to allocation. The negative churn reaction seems to be uniformly pervading all quarters at the micro level and the same phenomenon was evident at the supplier's level (macrobehavior model (2)-(5)) as well. Figure 2a and Figure 2b illustrate the high correlation between that churn vectors (lagged churn series). Moreover, the above results show that the highly correlated churn vectors are, in fact, consistent with customer responses (in at least 50% of the customers) to capacity rationing.

The *AR(1)* coefficients exhibit a pattern that mirrors the macrobehavior model, i.e., they increase in magnitude from $Cv^{-}q1(t)$ (0.2294) to $Cv^{-}q3(t)$ (0.4076) before dropping to 0.2655 in $Cv^{-}q4(t)$. The strong mean reversion property of the negative churn at the micro level is confirmed and is reflected in the macrobehavior at the supplier level as well.

Summary of micromotives analysis

To summarize the results of the micromotives analysis, we find strong statistical associations between customers' orders, positive and negative forecast churn, and prior-period allocation ($alloc_{t-1}$). Roughly 46% of instances where CUST-P/N pairs are rationed are followed by an order increase in the following period (these increases may be initiated by either customers or the supplier). While these order increases might be considered normal effects of rationing and backlogging, rationing also leads to significant forecast increases in future periods for CUST-P/N pairs that remain with the supplier going forward. Other important responses to rationing include leaving the supplier and batching of future forecasts. Whether these responses are customer driven or supplier driven, they indicate strong interdependencies between rationing and follow-up actions by the supplier and customers that have important implications for capacity management and supplier profitability.

9. Managerial Discussion

The results of the econometric model and its estimation are given in Figure 5a. These results strongly suggest that observed macrobehavior/supplier manufacturing facility level information distortions in a semiconductor supply chain are statistically associated with the mismatch between supply and demand at the micro level (CUST-P/N level, but dictated more by P/N than CUST). Extant literature in supply chain management points to demand uncertainty as a cause of correlated forecasts, the bullwhip effect and forecast batching. Building on the earlier theoretical results of Cachon and Lariviere, the results of this paper suggest that these effects may be reinforced by the supplier's allocation policy in interaction with underlying demand properties.

Demand Visibility: In a build-to-forecast manufacturing facility producing custom products with long production lead times, demand visibility is a difficult issue. The supplier bears the cost of overage and hence may use buffer inventory to smooth his production. Greater underlying demand uncertainty (variability at the customer level) and longer production lead times require larger amounts of buffer inventory to minimize total supply chain costs. In a large manufacturing facility (such as the one studied in this paper), the supplier depends on customers to use scientific forecasting methods (Time Series, MMFE, etc.), investments in information systems and coordinated downstream supply chains to gather information at least six months prior to production. This noisy forecast information is fed to the upper echelons of the supply chain for capacity planning and Sales & Operations Planning (S&OP) processes. The results of excess churn in this forecast information can therefore be very costly in terms of capacity utilization and profitability.

Supply Uncertainty: While demand uncertainty affects the supplier directly, the supplier's uncertain actions may cause customers to speculate about supply and hence providing noisy forecasts to the supplier. This inevitable feedback loop of allocation policy and forecasting is verified using the micromotives/macrobehavior model. We find that supply uncertainty is often related to the allocation policy, but is nevertheless a consequence of the supplier's tendency to delay capacity commitments to a later stage in the forecasting horizon.

Mismatch consequences: Our analysis reveals that the mismatch between supply and demand at the customer-P/N level causes four different results– positive and negative churn, bullwhipping, forecast batching, and contagion. One of the more interesting aspects of our findings is the tendency, whether customer driven or supplier driven, to batch forecasts around the beginning and end of quarters, a tendency that confounds the supplier's ability to conduct temporal smoothing of production. Moreover, a comparison of the mean reverting properties (see *AR(1)* coefficients in Table 3) of the customer requests and churn suggests that churn is more volatile than requests. As one symptom of this, supplier manufacturing facility level forecasts for the twelve months preceding an order are between 22%-40% (depending on the month) more volatile than the associated final requests/orders made by customers. We also found that customers tend to churn their future forecasts positively in response to allocation. The result is that forecast churn interacts with rationing to obfuscate the supplier's view of downstream temporal demand, resulting in a higher buffer stock and lower capacity utilization for the supplier. In a build to forecast system where forecasts six months prior to order commitment are used to pass on orders to upper echelons, these combined effects present a serious threat of bullwhipping.

We also find evidence for this churning of forecasts to be occurring in tandem across large pools of CUST-P/N(s), giving rise to inflated demands for capacity at precisely the same period that the supplier finds himself short of capacity. We refer to this cross-sectional (CUST-P/N) phenomenon of positively correlated forecasts as *contagion*. Clearly, such positive correlation erodes the benefits arising from risk pooling of CUST-P/N orders.

Managing Supply-Demand Mismatches

We have thus far taken an investigative approach, seeking symptoms of information and forecast distortion in the semiconductor supply chain. In seeking ways to mitigate forecast distortions, the appropriate starting point is to investigate the root cause of the supply-demand mismatches in the supply chain. Prescriptions should therefore be directed to treating this underlying root cause rather than ad hoc solutions to suppress the malignant symptoms of churn, system exiting, batching, bullwhipping and contagion. Such prescriptions will certainly include better linking of customers and suppliers through CPFR (collaborative planning, forecasting and replenishment) practices, and may include modified allocation policies and innovations in contracting such as options-based contracts. Let us consider these possibilities.

CPFR Practices: CPFR is an initiative launched by this supplier with its customers whereby the arm's length nature of relationships are transitioned to more long-term relationship based contracts. In addition to providing better information between suppliers and customers, the relationship-specific investments required in CPFR reinforce loyalty to the relationship. Aviv (2001, 2002) quantifies the value of CPFR on supply chain performance, where this value is based on better information, reduced manufacturing lead times and increased supplier production flexibility to monthly variations in orders. With a more reactive and flexible supplier, the buyer's incentive to batch forecasts is reduced. CPFR initiatives targeted at increasing forecast accuracy and reducing the information asymmetry between customers and the supplier encompass the gamut of forecast practices such as joint IT investments and making judgmental forecasts more accurate. Accurate forecasts of end user demand naturally address the issue of supply-demand uncertainty.

Embedded Optionality in contracts

In a manufacturing environment faced with contagion where risk pooling is rendered ineffective, we envision the use of tailored options-based contracts to develop a diversified demand-forecast portfolio. (See de Albeniz and Simchi-Levi, 2005a.) The notion of diversification through risk pooling is essential to capacity planning models in build to forecast environments. As de Albeniz and Simchi-Levi (2005) show, diversification can be obtained through embedding options within contracts even though the underlying demand might not provide good diversification possibilities. The use of options at this supplier was through pricing discounts for early commitments (the customer gives up some flexibility in commitment dates in return for a price discount). Such options could be useful in controlling contagion through obtaining fixed, early commitments, which could also help in mitigating forecast churn, as described below.

Since, the economic costs of churn affect the supplier and the upper echelons of the supply chain, conceptually it seems feasible to pass the costs of this churn through to the customer by contracting on forecasts of the underlying capacity instead of on the capacity (as is the norm). Our discussions with representatives from the semiconductor industry suggested that pricing policies packaged as penalty structures run against the spirit of collaborative efforts and hence equivalent transfer pricing mechanisms which reward "non-churn" would be more acceptable to customers. By contracting on forecasts of the capacity needs instead of capacity itself, the customer would thus be provided with incentives to pay greater attention to forecasts lest his contractual forecast costs burgeon. For example, by rewarding customers who place firm orders earlier in the forecasting cycle with higher priority in the capacity allocation process or by providing pricing discounts, the supplier encourages shorter forecasting cycles and consequently lower churn. Ferguson (2003) and Ferguson et al. (2004) provide a rigorous analysis of the buyer's decision to commit and update forecasts.

Allocation Policies

While our empirical analysis focused primarily on allocation policies, CPFR mechanisms have tended to focus entirely on relationship-based contracts and explicit rules for resolving demand-supply mismatches. This is in contrast to the semiconductor industry's proclivity toward undisclosed allocation policies and the resulting arm's length nature of resolving demand-supply mismatches. For customers and part numbers which are not amenable to the direct resolution methods of CPFR and relationship-based contracting, the supplier could nonetheless significantly reduce the mismatch between supply and demand by expanding the dimension of allocation policies to time and commitment based policies. To illustrate the point; consider two customers, one that commits early in the forecasting process (at $t=-11$) and one that commits late in the forecasting process ($t=-6$). By giving priority to the early commitment forecast, the supplier has the freedom to plan his capacity well in advance (thus reducing churn) and communicate his orders more accurately to the upper echelons (reduce bullwhip). Since the two customers commit at different times, the contagion effect (by definition) vanishes.

Allocation policies could also be modified to isolate customers with highly correlated demand profiles. By dedicating capacity to sub groups of customers such that within-group heterogeneity is maximized while between-group homogeneity is maintained, the supplier can effectively partition the underlying capacity commitments into sub groups. The notion of maintaining between-group homogeneity and within-group heterogeneity is meant to diversify within-group demand volatility, but still gain economies of scale due to the between-group homogeneity.

The above represent a few areas where future empirical work has the potential for improving supplier profitability and customer satisfaction in the semiconductor industry. Other areas include the development of appropriate metrics for monitoring the performance and cost consequences of measures directed to mitigate churn and contagion and refinements in relationship-based contracting to share the gains from such improved performance. Certainly, the diversity of behavioral responses observed in this study suggests that further development of supporting theory will need to be informed by the results of detailed empirical research. The required analysis and experimentation should provide a good opportunity for collaboration between the research community and industry.

References

Aviv, Y. 2001. The Effect of Collaborative Forecasting on Supply Chain Performance. *Management Science* **47** (10). 1326-1343.

Aviv, Y., Federgruen, A. 2001. Design for Postponement: A Comprehensive Characterization of its Benefits Under Unknown Demand Distributions. *Operations Research* **49** (4). 578-598

Aviv, Y. 2002. Gaining Benefits from Joint Forecasting and Replenishment Processes: The Case of Auto-correlated Demand. *Manufacturing & Service Operations Management* **4** (1). 55-74.

Cachon, G. and M. Lariviere. 1999a. An equilibrium analysis of linear and proportional allocation of scarce capacity. *IIE Transactions*. **31** (9). 835-850.

Cachon, G., M. Lariviere. 1999b. Capacity choice and allocation: strategic behavior and supply chain performance. *Management Science*. **45** (8) 1091-1108

Cachon, G., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* **46**(8). 1032-1048.

Cattani, K., Hausman, W. 2000. Why are Forecast Updates Often So Disappointing. *Manufacturing and Service Operations Management*. **2(**2). 119-127

Chen, F., Drezner. Z., Ryan, K.J., Simchi-Levi, D. 2000. Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information. *Management Science*. **46** (3). 436-443

Croson, R.T.A., K. Donohue. 2006. Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information. Forthcoming, *Management Science*.

de Albeniz, M.V., D. Simchi-Levi (2005), Mean-Variance Trade-offs in Supply Contracts. To appear in *Naval Research Logistics.*

de Albeniz, M.V., D. Simchi-Levi (2005), A Portfolio Approach for Procurement Contracts. *Production and Operations Management.* **14.** 90-114.

Ferguson, E. M. 2003. When to Commit in a Serial Supply Chain with Forecast Updating. *Naval Research Logistics*. **50**. 917-936

Ferguson, E. M., DeCroix, A.G., Zipkin, H. P., 2004 Commitment Decisions with Partial Information Updating. Working Paper, Dupree College of Management, Georgia Institute of Technology

Forrester, J.W. 1958. Industrial Dynamics: A Major Breakthrough for decision makers. *Harvard Business Review*. July-Aug 1958.

Gaur, V., Avi Giloni, Sridhar Seshadri. 2005. Information Sharing in a Supply Chain under ARMA Demand. *Management Science.* **51**(6).

Gavirneni, S., Kapuscinki, R., Tayur, S. 1999. Value of Information in Capacitated Supply Chains. *Management Science* **45**(1), 16-24.

Graves, S.C., D. B. Kletter ,W. B. Hetzel. 1998. A Dynamic Model for Requirements Planning with Application to Supply Chain Optimization, *Operations Research*, May-June 1998, **46**(3). S35-S49.

Intel Press Release. 2005. Innovation More Important Than Ever in a Platform Era. *Intel Developer Forum*, San Francisco.

Kaminsky, P., Swaminathan, M, J. 2001. Utilizing Forecast Band Refinement for Capacitated Production Planning. *Manufacturing and Service Operations Management*. **3** (1). 68-81

Hausman, Warren. 1969. Sequential Decision Problems: A Model to Exploit Existing Forecasters. *Management Science*. **16**(2). B93-B111

Heath, D.C., P.L.Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*. **26**(3). 17-30.

Heching, Aliza and Alan King. 2005. Supplier Managed Inventory for Custom Manufactured Items with Long Lead Times. Working Paper, IBM Research (September).

Kleindorfer, Paul R. and D.J. Wu, 2003. "Integrating Long-term and Short-term Contracting via Business-to-Business Exchanges for Capital-Intensive Industries", *Management Science.* **49** (11) 1597-1615.

Lee, H., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science*. **43** (4). 546-58.

Maddala, G.S. 1992. *Introduction to Econometrics.* 2nd edition. Macmillan Publishing Company.

Milner, J., Kouvelis, P. 2002 On the Complementary Value of Accurate Demand Information and Production and Supplier Flexibility. *Manufacturing and Service Operations Management*. **4** (2). 99-113

Miyaoka, J., Hausman., W. 2004. How a Base Stock Policy Using “Stale Forecasts Provides Supply Chain Benefits. *Manufacturing and Service Operations Management*. **6** (2) 149-162.

Terwiesch, C., Z. J. Ren, T. Ho, and M. Cohen. 2004. An Empirical Analysis of Forecast Sharing in the Semiconductor Equipment Industry. *Management Science*, **51** (2). 208-220.

Schelling, Thomas 1978. *Micro Motives and Macro Behavior*. W.W. Norton & Company

Taylor, T.A. and Plambeck, E.L. (2003) Supply Chain Relationships and Contracts: The Impact of Repeated Interaction on Capacity Investment and Procurement, working paper, Stanford GSB.

Toktay, L.B. and L.M. Wein, "Analysis of a Forecasting-Production-Inventory System with Stationary Demand," *Management Science* **47** (9). 1268-1281.

Watson, Noel, and Yu-Sheng Zheng, 2005. "Adverse Effects of the Over-estimation of the Permanence of New Demand Levels in Subjective and Quantitative Decision Making." Working Paper, OPIM, The Wharton School.

Fig 1: Timeline of buyer-supplier interactions

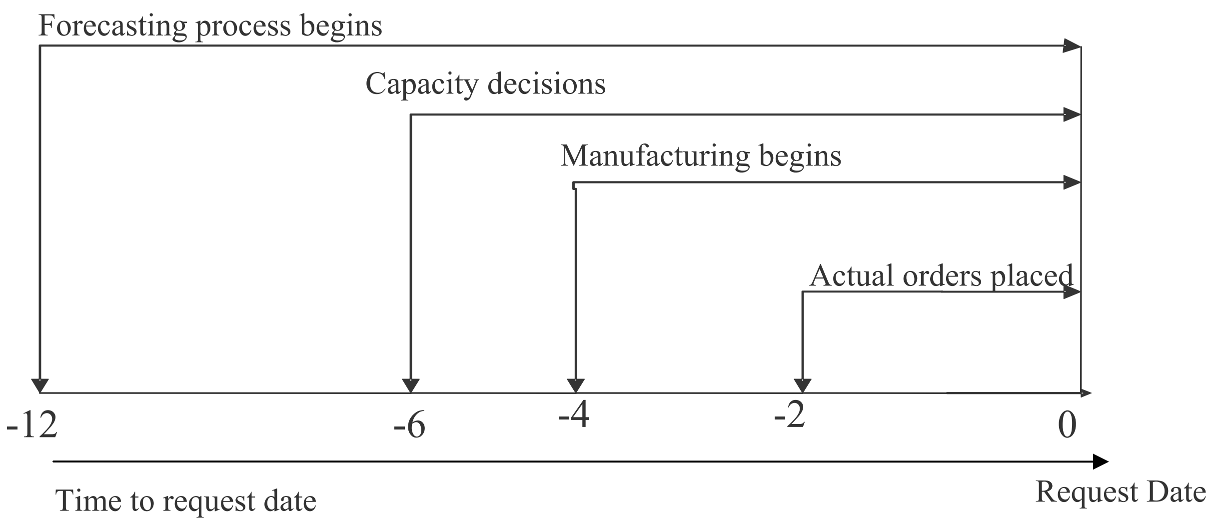

Fig 2: System Dynamics illustrating customer's actions before and after placing an order

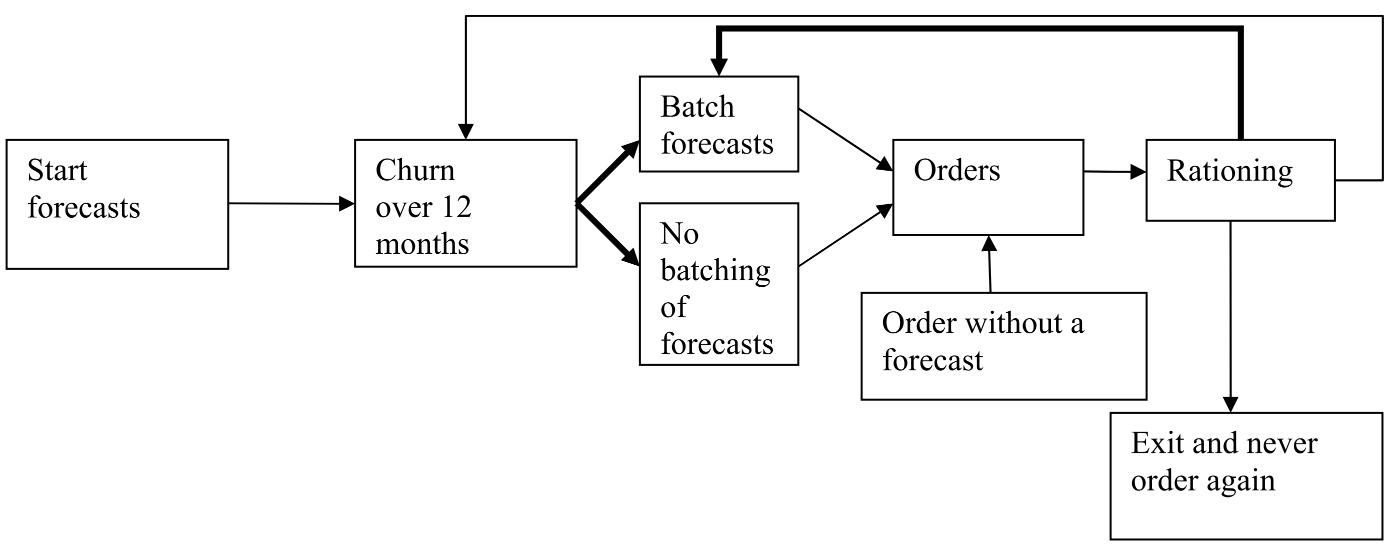

Table 1: Summary Statistics for the entire database

Variable	Count	Min	Max	Mean	Standard deviation	Description
TIME	60	01/2000	12/2004	-	-	Requested Delivery date
CUST	2565	-	-	-	-	Unique Customer ID
P/N	11868	-	-	-	-	Unique part number
CUST-P/N	28800	-	-	-	-	Unique Cust-p/n combination
SPOT	13575	0	1	-	-	Indicates if CustId-PartNo arrived without forecast
SYSEXIT	20523	0	1	-	-	Indicates if CustId-P/n exited and never transacted again
All statistics below are for a unit of the cross section (cust-p/n)						
START	-	-12	0	-7	3.7	# of months prior to requested date when forecast start
FINISH	-	-12	0	-2	3.5	# of months prior to requested date when forecasts end
$Qf_{t-12,t}$	-	0	2400000	469	20179	Forecast 12 months ahead of requested delivery date
$Qf_{t-11,t}$	-	0	24594500	4913	101128	Forecast 11 months ahead of requested delivery date
$Qf_{t-10,t}$	-	0	12000000	6077	93274	Forecast 10 months ahead of requested delivery date
$Qf_{t-9,t}$	-	0	9000000	6734	96369	Forecast 9 months ahead of requested delivery date
$Qf_{t-8,t}$	-	0	6716000	6993	97625	Forecast 8 months ahead of requested delivery date
$Qf_{t-7,t}$	-	0	23000000	7437	111753	Forecast 7 months ahead of requested delivery date
$Qf_{t-6,t}$	-	0	7582500	7480	102243	Forecast 6 months ahead of requested delivery date
$Qf_{t-5,t}$	-	0	8278000	7517	102815	Forecast 5 months ahead of requested delivery date
$Qf_{t-4,t}$	-	0	19862480	7450	107058	Forecast 4 months ahead of requested delivery date
$Qf_{t-3,t}$	-	0	7564265	7375	105507	Forecast 3 months ahead of requested delivery date
$Qf_{t-2,t}$	-	0	8897000	7209	108096	Forecast 2 months ahead of requested delivery date
$Qf_{t-1,t}$	-	0	9865000	7348	115539	Forecast 1 months ahead of requested delivery date
$Qf_{t,t}$	-	0	8717000	7150	113707	Last forecast made prior to making order
Qr_t	-	0	8739902	6341	103675	Quantity requested by the customer
Qc_t	-	0	9612500	6412	104947	Quantity committed to the customer by the supplier

Fig 3a: Forecast vectors of length 5 for months 1,2,3,4,5.

$$\begin{pmatrix} Q_{f11} & Q_{f22} & Q_{f33} & Q_{f44} & Q_{f55} \\ Q_{f12} & Q_{f23} & Q_{f34} & Q_{f45} & Q_{f56} \\ Q_{f13} & Q_{f24} & Q_{f35} & Q_{f46} & Q_{f57} \\ Q_{f14} & Q_{f25} & Q_{f36} & Q_{f47} & Q_{f58} \\ Q_{f15} & Q_{f26} & Q_{f37} & Q_{f48} & Q_{f59} \end{pmatrix}$$

Fig 3b: Churn quantities for each month.

$$\left(\begin{array}{ccccccccc} Q_{f11} & & Q_{f22} & & Q_{f33} & & Q_{f44} & & Q_{f55} \\ & \mathrm{C}_{12} & & \mathrm{C}_{23} & & \mathrm{C}_{34} & & \mathrm{C}_{45} & \\ Q_{f12} & & Q_{f23} & & Q_{f34} & & Q_{f45} & & Q_{f56} \\ & \mathrm{C}_{13} & & \mathrm{C}_{24} & & \mathrm{C}_{35} & & \mathrm{C}_{46} & \\ Q_{f13} & & Q_{f24} & & Q_{f35} & & Q_{f46} & & Q_{f57} \\ & \mathrm{C}_{14} & & \mathrm{C}_{25} & & \mathrm{C}_{36} & & \mathrm{C}_{47} & \\ Q_{f14} & & Q_{f25} & & Q_{f36} & & Q_{f47} & & Q_{f58} \\ & \mathrm{C}_{15} & & \mathrm{C}_{26} & & \mathrm{C}_{37} & & \mathrm{C}_{48} & \\ Q_{f15} & & Q_{f26} & & Q_{f37} & & Q_{f48} & & Q_{f59} \end{array}\right)$$

Fig 3c: Positive and Negative Churn obtained from the churn matrix

$$\begin{pmatrix} \mathrm{C}_{12} & \mathrm{C}_{23} & \mathrm{C}_{34} & \mathrm{C}_{45} \\ \mathrm{C}_{13} & \mathrm{C}_{24} & \mathrm{C}_{35} & \mathrm{C}_{46} \\ \mathrm{C}_{14} & \mathrm{C}_{25} & \mathrm{C}_{36} & \mathrm{C}_{47} \\ \mathrm{C}_{15} & \mathrm{C}_{26} & \mathrm{C}_{37} & \mathrm{C}_{48} \end{pmatrix} = \begin{pmatrix} \mathrm{C}^{+}_{12} & \mathrm{C}^{+}_{23} & \mathrm{C}^{+}_{34} & \mathrm{C}^{+}_{45} \\ \mathrm{C}^{+}_{13} & \mathrm{C}^{+}_{24} & \mathrm{C}^{+}_{35} & \mathrm{C}^{+}_{46} \\ \mathrm{C}^{+}_{14} & \mathrm{C}^{+}_{25} & \mathrm{C}^{+}_{36} & \mathrm{C}^{+}_{47} \\ \mathrm{C}^{+}_{15} & \mathrm{C}^{+}_{26} & \mathrm{C}^{+}_{37} & \mathrm{C}^{+}_{48} \end{pmatrix} + \begin{pmatrix} \mathrm{C}^{-}_{12} & \mathrm{C}^{-}_{23} & \mathrm{C}^{-}_{34} & \mathrm{C}^{-}_{45} \\ \mathrm{C}^{-}_{13} & \mathrm{C}^{-}_{24} & \mathrm{C}^{-}_{35} & \mathrm{C}^{-}_{46} \\ \mathrm{C}^{-}_{14} & \mathrm{C}^{-}_{25} & \mathrm{C}^{-}_{36} & \mathrm{C}^{-}_{47} \\ \mathrm{C}^{-}_{15} & \mathrm{C}^{-}_{26} & \mathrm{C}^{-}_{37} & \mathrm{C}^{-}_{48} \end{pmatrix}$$

Fig 4a: Negative churn experienced by facility in each quarter, i.e., ***quarterly*** churn series ***for*** a delivery date t

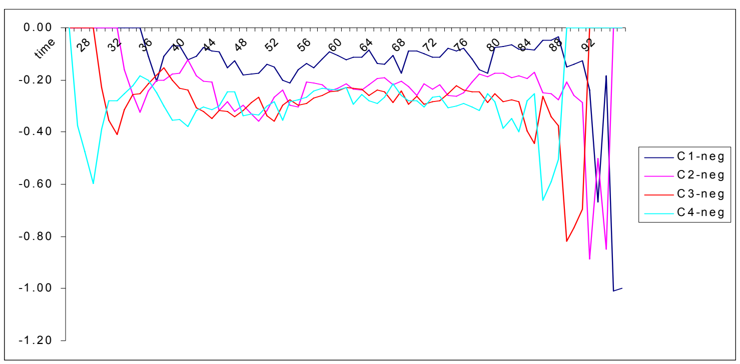

Fig 4b: Lagged quarterly churn experienced by a facility, i.e., ***quarterly*** churn vector ***at*** a given time

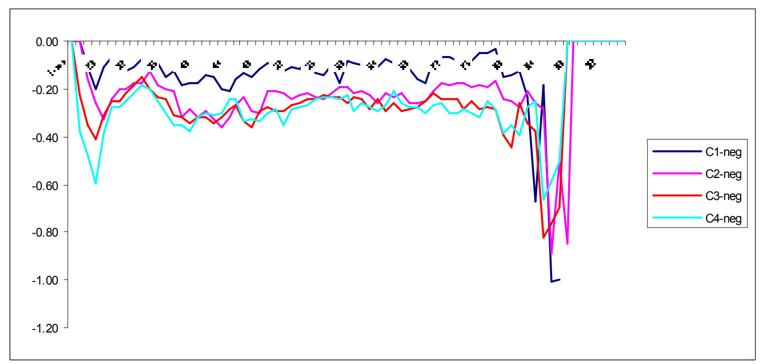

Table 2a: Correlation between quarterly churn *series* components i.e. quarterly churn ***for*** time t

	$Cs^{-}q1(t)$	$Cs^{-}q2(t)$	$Cs^{-}q3(t)$	$Cs^{-}q4(t)$
$Cs^{-}q1(t)$	1.00	0.81	0.24	0.03
$Cs^{-}q2(t)$	0.81	1.00	0.39	-0.10
$Cs^{-}q3(t)$	0.24	0.39	1.00	0.49
$Cs^{-}q4(t)$	0.03	-0.10	0.49	1.00

Table 2b: Correlation between quarterly churn *vector* components i.e. quarterly churn ***at*** time t

	$Cv^{-}q1(t)$	$Cvs^{-}q2(t)$	$Cv^{-}q3(t)$	$Cv^{-}q4(t)$
$Cv^{-}q1(t)$	1.00	0.86	0.71	0.43
$Cv^{-}q2(t)$	0.86	1.00	0.85	0.65
$Cv^{-}q3(t)$	0.71	0.85	1.00	0.78
$Cv^{-}q4(t)$	0.43	0.65	0.78	1.00

Table 3: LSDV estimator for Macro-behavior Regression Model

	$Qr(t)$	$Cv^{+}q1(t)$	$Cv^{+}q2(t)$	$Cv^{+}q3(t)$	$Cv^{+}q4(t)$	$Cv^{-}q1(t)$	$Cv^{-}q2(t)$	$Cv^{-}q3(t)$	$Cv^{-}q4(t)$	*sysexits(t)*
$Qr(t-1)$	0.8037*** (0.1254)									
$Cv^{+}q1(t-1)$		0.6168*** (0.0924)								
$Cv^{+}q2(t-1)$			0.6313*** (0.891)							
$Cv^{+}q3(t-1)$				0.2464** (0.1211)						
$Cv^{+}q4(t-1)$					0.0609 (0.1419)					
$Cv^{-}q1(t-1)$						0.4037*** (0.1201)				
$Cv^{-}q2(t-1)$							0.622*** (0.1266)			
$Cv^{-}q3(t-1)$								0.7306*** (0.0828)		

$Cv^{-}q4(t-1)$									0.5627*** (0.1176)	
sysexits(t-1)										0.6399*** (0.1189)
Falloc(t-1)	4.4e7*** (1.2e7)	0.0431** (0.0191)	0.1087*** (0.0284)	0.2394*** (0.0595)	0.3897*** (0.0771)	-0.1952** (0.0910)	-0.2723* (0.1403)	-0.24** (0.1001)	-0.3691** (0.1560)	0.6479*** (0.1205)
DUMMYM1	-1.4e7*** (2.3e6)	0.0035 (0.0034)	0.0025 (0.0044)	0.0277** (0.0111)	0.0132 (0.0149)	-0.0473** (0.022)	-0.072** (0.032)	-0.055*** (0.0199)	-0.1288*** (0.0304)	-188.57*** (46.04)
DUMMYM2	-4.823e6** (1.8e6)	0.0095** (0.0039)	0.0019 (0.0047)	0.0162 (0.0128)	0.0015 (0.0169)	-0.0441* (0.026)	-0.018 (0.034)	-0.0088 (0.0213)	-0.0242 (0.0334)	-81.70*** (38.84)
R^2 p-value	0.9502 0	0.9199 0	0.9627 0	0.8509 0	0.7986 0	0.7286 0	0.8469 0	0.9502 0	0.9055 0	88.15% 0

Note: Standard errors are in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% levels, respectively

Table 4a: Embedded Stochastic coefficients VAR estimator for Micromotives model of buyer-customer interactions (*Qr-alloc* simultaneous estimator).

	CustId Model 1		PartNo Model 2		CustId-PartNo Model 3	
	Qr(t)	*alloc(t)*	*Qr(t)*	*alloc(t)*	*Qr(t)*	*alloc(t)*
Fixed Effects Intercept		-0.007*** (0.0022)		-0.0066*** (0.0022)		-0.0051** (0.0021)
Qr(t-1)	0.099*** (0.003)	0.1855*** (0.0074)	0.1006*** (0.003)	0.1955*** (0.0054)	0.1024*** (0.0031)	0.1843*** (0.0041)
alloc(t-1)	0.0278*** (0.006)	-0.042*** (0.0065)	0.0227*** (0.004)	-0.0482*** (0.0046)	0.0214*** (0.0039)	-0.046*** (0.0041)
Random Effect Intercept		0.0035		0.0048		0.004
Qr(t-1)		0.149		0.2293		0.273
alloc(t-1)	0.115	0.115	0.163	0.1612	0.1901	0.202
AIC	317354.3	315900.6	316970.7	314524.5	316449.7	313034.9

BIC	317393.4	315998.5	317009.8	314622.4	316488.8	313132.8
LogLik	-158673.1	-157940.3	-158481.3	-157252.2	-158220.8	-156507.4

Note: Standard errors are in paranthesis. ***, **, * denote significance at the 1%, 5% and 10% level respectively

Table 4b: Embedded Stochastic coefficients REML estimator of customer's positive churn (Micromotives) behavior.

	$Cv^{+}q1(t)$			$Cv^{+}q2(t)$		
	CUST	P/N	CUST-P/N	CUST	P/N	CUST-P/N
Fixed Effects						
$Cv^{+}q1(t\text{-}1)$	0.270***	0.1932***	0.1993***			
	(0.0042)	(0.0043)	(0.0043)			
$Cv^{+}q2(t\text{-}1)$				0.3546***	0.3546***	0.3546***
				(0.0035)	(0.0035)	(0.0035)
alloc(t-1)	-0.0001	-0.0129**	-0.012***	-0.0004*	-0.00054*	-0.00054*
	(0.0005)	(0.005)	(0.0039)	(0.0002)	(0.0003)	(0.0003)
Random Effect						
alloc(t-1)	0.00162	0.1808	0.1571	0.0004	0.0007	0.0007
AIC	14464.31	12537.17	12801.34	34239.7	34237.62	34237.64
BIC	14499.8	12572.66	12836.83	34276.09	34274.02	34274.04
LogLik	-7228.155	-6264.583	-6396.672	-17115.85	-17114.81	-17114.82

	$Cv^{+}q3(t)$			$Cv^{+}q4(t)$		
	CUST	P/N	CUST-P/N	CUST	P/N	CUST-P/N
Fixed Effects						
$Cv^{+}q3(t\text{-}1)$	0.2754***	0.2742***	0.2736***			
	(0.0034)	(0.0033)	(0.0034)			
$Cv^{+}q4(t\text{-}1)$				0.1172***	0.1161***	0.1154***
				(0.0034)	(0.0034)	(0.0034)
alloc(t-1)	-0.0052***	-0.02***	-0.022***	0.0002*	-0.0091***	-0.01***
	(0.0011)	(0.002)	(0.003)	(0.0000)	(0.002)	(0.0021)
Random Effect						
alloc(t-1)	0.0078	0.0553	0.0659	0.0000	0.0316	0.044

AIC	55796.66	55593.43	55531.59	49351.3	49373.18	49366.23
BIC	55833.83	55630.6	55568.76	49388.47	49410.36	49403.4
LogLik	-27894.33	-27792.71	-27761.79	-24671.65	-24682.59	-24679.11

Note: Standard errors are in paranthesis. ***, **, * denote significance at the 1%, 5% and 10% level respectively

Table 4c: Embedded Stochastic coefficients REML estimator of customer's negative churn (Micromotives) behavior.

	$Cv^{-}q1(t)$			$Cv^{-}q2(t)$		
	CUST	P/N	CUST-P/N	CUST	P/N	CUST-P/N
Fixed Effects						
$Cv^{-}q1(t-1)$	0.2469***	0.2294***	0.2287***			
	(0.0046)	(0.0045)	(0.0045)			
$Cv^{-}q2(t-1)$				0.2456***	0.2328***	0.2276***
				(0.0037)	(0.0036)	(0.0036)
alloc(t-1)	0.0001	0.0212***	0.0208***	0.0001	0.0136***	0.01191***
	(0.0001)	(0.0063)	(0.0048)	(0.00001)	(0.0042)	(0.0032)
Random Effect						
alloc(t-1)	5.1e-7	0.2402	0.2113	6.4e-7	0.1774	0.1747
AIC	26304.41	22386.47	22669.71	6773.368	2103.181	1103.359
BIC	26339.9	22421.96	22705.2	6810.898	2140.711	1140.889
LogLik	-13148.2	-11189.23	-11330.85	-3382.684	-1047.59	-547.6796

	$Cv^{-}q3(t)$			$Cv^{-}q4(t)$		
	CUST	P/N	CUST-P/N	CUST	P/N	CUST-P/N
Fixed Effects						
$Cv^{-}q3(t-1)$	0.2658***	0.2665***	0.2628***			
	(0.0039)	(0.0038)	(0.0039)			
$Cv^{-}q4(t-1)$				0.2653***	0.2665***	0.2628***
				(0.0039)	(0.0038)	(0.0039)
alloc(t-1)	0.0451***	0.075***	0.0953***	0.0451***	0.0749***	0.0953***
	(0.001)	(0.0212)	(0.001)	(0.0049)	(0.0212)	(0.0104)
Random Effect						

alloc(t-1)	0.063	1.015	0.5640	0.063	1.011	0.564
AIC	82114.22	80928.8	82131.68	82114.22	80928.8	82131.68
BIC	82150.78	80965.36	82168.23	82150.78	80965.36	82168.23
LogLik	-41053.11	-40460.4	-41061.84	-41053.11	-40460.4	-41061.84

Note: Standard errors are in paranthesis. ***, **, * denote significance at the 1%, 5% and 10% level respectively

Fig 5a: The consequences of supply-demand mismatch(es)