Data Set: 187 Weeks of Customer Forecasts and Orders for Microprocessors from Intel Corporation

February 16, 2020

Matthew P. Manary

Data Center Group, Intel Corporation, Hillsboro, Oregon 97124 `matthew.manary@intel.com`

Sean P. Willems

Haslam College of Business, University of Tennessee, Knoxville, TN 37996 `willems@utk.edu`

Abstract

Problem definition: This data set contains 187 consecutive weeks of Intel microprocessor demand information for all five distribution centers in one of its five sales geographies. For every stock keeping unit (SKU) at every location, the weekly forecasted demand and actual customer orders are provided, as well as the SKU's average selling price category. This data is provided by week and by distribution center, producing 26,114 records in total. **Relevance**: The 86 SKUs in the data set span five product generations. It provides years of product evolution across generations and price points. **Methodology**: As a data set paper, its purpose is to provide interesting and rich real-world data for researchers developing forecasting, inventory, pricing, and product assortment models. **Results**: The data set demonstrates the presence of significant forecast bias, heterogeneity of forecast errors between distribution centers, generational differences, product life cycles, and pricing dynamics. **Managerial Implications**: This data set provides access to a rich pricing and sales setting from a major corporation that has not been made available before.

1 Introduction

Intel's participation in the 2018 Edelman Competition served as the catalyst to create this data set paper. Manary et al. (2019) document Intel's ten-year journey to implement a fully automated inventory planning system requiring no planner involvement. The two most significant technical challenges the team had to overcome were resolving forecast bias and mapping each newly introduced microprocessor to an existing order stream. While this data set helps put the problem setting from Manary et al. (2019) in perspective, it also gives researchers the opportunity to test forecasting, inventory, pricing, and product assortment models against real-world data.

The 26,114 records in the data set represent weekly-level unit forecasts, customer orders, and price bands for 86 stock keeping units (SKUs) spanning five generations in five distribution centers across 187 consecutive weeks. It is reasonable to think of all 86 SKUs as desktop microprocessors or notebook microprocessors or server microprocessors where all the SKUs may be considered generally substitutable with one another at varying levels of price and performance.

This paper complements two *M&SOM* data set papers that provide insight into real-world supply chain issues. Willems (2008) documents 38 supply chains created by analysts modeling multi-echelon inventory problems. Acimovic et al. (2019) provides 182 weeks of order information in Dell's business-to-consumer market for make-to-stock computers. Intel's data set is focused on the business-to-business market for microprocessors; Intel's problem can be thought of as occurring at one echelon upstream of Dell in the computer supply chain. At this upstream echelon, hardware OEMs like Apple, Dell, and Lenovo are Intel's customers.

Acimovic et al. (2019) share the actual calendar weeks in which orders were placed for an entire sales geography, but round and scale the order amounts to maintain company confidentiality. We maintain confidentiality by not providing the specific weeks the data set represents, instead noting the timeframe is between 2010 and 2017 and overlaps with Acimovic et al. (2019). The benefit is that we can share significantly more information per week, beyond just customer orders. For one of Intel's five sales geographies, we provide 187 consecutive weeks of customer orders as well as demand forecasts, product offering, product generation, and selling price category for every SKU by distribution center by week. To anonymize the unit quantities, we multiply all forecast and order data by a common scalar that is the same across SKUs. If this was not done, the volumes could reveal which Intel division the data came from. By applying the same scalar across all the

data, all the relative errors are identical to the original data.

Section 2 describes the data. Section 3 provides summary statistics for the data to establish a baseline understanding of the data. Section 4 points out some facets of the data Intel finds interesting to let researchers see how Intel thinks about this data. Section 5 concludes the paper while Section 6 explains how to access the data.

2   Explaining The Data

There are eight columns and 26,114 records in the data set. Table 1 shows the first three records.

Distribution Center	Product Offering	Generation	SKU	ASP Group	Week	Forecasted Demand	Customer Orders
ALPHA	A	2	SKU-A-2	1	1	8949	11146
ALPHA	B	2	SKU-B-2	1	1	11146	3503
ALPHA	C	1	SKU-C-1	3	1	1274	5892

Table 1: Three of the 26,114 records in the data set.

Distribution Center represents an Intel-owned facility maintaining finished goods inventory to directly service customer demand. A customer could be an electronics distributor that resells the SKU or an original equipment manufacturer buying direct from Intel to build a system or an Intel-operated VMI hub that, as described in Manary et al. (2019), is located in a customer's manufacturing site. Five distribution centers, labeled ALPHA, BETA, GAMMA, DELTA, and EPSILON, comprise all of the distribution centers in the sales geography. Each distribution center serves more than one customer. Some customers may operate multiple locations in the region but each customer location orders from a specific distribution center. Distribution centers are geographically dispersed from one another and Intel operates each distribution center as if it's demand process is independent from the other distribution centers. Transshipments between distribution centers are possible but very rarely occur.

Product Offering reflects Intel's intent to serve a particular market. Product offering can be thought of as representing both model and good-better-best market positioning. For example, Table 2 presents four Dell notebooks that are each equipped with different Intel processors, with "i5" being generally better than "i3", and an 8265U being a different model than an 8130U. Each combination of model and market position is assigned its own unique alpha character. The 25 alpha characters denoting product offering are not rank ordered; e.g., product offering "A" is not

necessarily superior to "B".

Generation is a numeric value defined by the microprocessor architecture and manufacturing process technology. The five generations in the data set are rank ordered with each higher number representing a more recent, and improved, generation. For example, in Table 2 there are processors represented from both the 7th process technology generation and 8th generation. If the 7th generation were the first generation of products in the data set then the 7th generation products would be designated as generation "1" in the data and the 8th generation processors would be shown as belonging to generation "2".

Dell Computer	Latitude 5424 Rugged BusinessLaptop $2,089.47 **$1,459.00** 7th Gen Intel® Core™ i3-7130U Processor Windows 10 Pro 64bit English 8GB, 2x4GB, 2400MHz DDR4 Non-ECC 2.5" 500GB 7200RPM SATA Hard Drive	Latitude 7390 $1,712.86 **$1,199.00** 8th Gen Intel® Core™ i3-8130U Windows 10 Pro 64bit English 4GB, 1x4GB, DDR4 2400MHz Memory M.2 128GB SATA Class 20 Solid State Drive	New Inspiron 15 3000 $709.99 **$659.99** 8th Generation Intel® Core™ i5-8265U Processor Windows 10 Pro 64-bit English 8GB, 1x8GB, DDR4, 2666MHz 1TB 5400 rpm 2.5" SATA Hard Drive	Vostro 15 5000 $990.00 **$689.00** 7th Generation Intel® Core™ i5-7200U Processor Windows 10 Pro 64-bit English 4GB, DDR4, 2400MHz; up to 32GB 1TB 5400 rpm Hard Drive
Hypothetical Intel SKU	SKU—Z—1	SKU—Z—2	SKU—BB—2	SKU—AA—1

Table 2: Example to map four Intel SKUs that represent three product offerings and two generations.

SKU is one of the 86 microprocessors in the data set defined by uniquely pairing a product offering and generation. Using the Dell Smart Selections product line studied in Acimovic et al. (2019), Table 2 demonstrates varying combinations of product offerings and process generations. For example, Dell's Latitude 5424 Rugged Business Laptop comes with a 7th generation i3-7130U processor, whereas the Latitude 7390 comes with an 8th generation i3-8130U. These two product offerings are the same (i.e. an "i3-x130U" product), with the difference being the process technology (7th vs. 8th generation). Hypothetically, if the i3-x130U microprocessor is denoted by product offering "Z" and we assume the 7th process technology was the earliest in the data set, then the 7th generation i3-7130U would appear in the data set as SKU-Z-1 whereas the 8th generation i3-8130U would appear as SKU-Z-2. Similarly, the Vostro 15 5000 laptop in Table 2 has a 7th

generation processor, but the product offering i5-7200U is different. If the i5-7200U microprocessor is mapped to product offering "AA", then the i5-7200U in the data set would be SKU-AA-1. In a similar fashion, the 8th generation i5-8265U in the New Inspiron 15 3000 laptop might be listed as SKU-BB-2.

While there are 125 possible pairings of product offering and generation, only 86 SKUs exist in the data set. Intel regularly adds new product offerings in each generation, and may also retire a certain product offering whose unique features are no longer demanded by the market.

Week represents an ordered calendar week. Week 1 is the earliest week in the data set and week 187 is the latest, or most recent. Our guiding principle in choosing the duration was to reflect a period of economic, competitive and production stability. The duration avoids significant economic swings like the global economic downturn from 2008 to 2010. The forecast errors are not a function of sales significantly misestimating market share nor are the customer orders inflated to compensate for production challenges. There were no significant production issues or backlogs in the region during this time frame. While company confidentiality prevents us from disclosing the exact calendar dates, we can say the data was taken from between 2010 and 2017 and overlaps with Acimovic et al. (2019). There are 10 SKUs, all from generation 3 (A, B, C, E, H, I, J, N, S, T) that experienced a full life-cycle within the 187 week period.

ASP Group is a collection of SKUs whose average selling price (ASP) is 15% higher than the average price of the ASP group one level down from it, and there are 22 such groups. These 22 prices are invariant across the 187 weeks. ASP group n in week 1 is the same 15% price band as ASP group n in week 187. The ASP groups are in rank order, such that every SKU in ASP group n is more expensive than any SKU in ASP group $n-1$. Therefore, the average price of SKUs in ASP group 22 are approximately 19 times more expensive than ASP group 1 ($1.15^{21} = 18.8$).

ASP group 20 is not present in this data set as no products from that ASP group were sold in this timeframe in this geography. As such, SKUs in ASP group 21 are on average 32% (1.15^2) more expensive than ASP group 19. In general, performance and features correlate with price such that SKUs from ASP group 22 are generally superior to products in ASP group 1. A SKU stays in the same ASP group through its life cycle but it is common for SKUs associated with a given product offering to change from generation to generation. For example, SKU-K-1 is in ASP group 10, SKU-K-2 is in ASP group 11, and SKU-K-3, SKU-K-4, and SKU-K-5 are in ASP group 12.

Forecasted Demand represents Intel's estimate of weekly orders in units for a given SKU at

a specific location made after customers submit their forecasts. Large customers, like original equipment manufacturers, provide a forecast for every SKU in a specific week at a time interval sufficient for the distribution center to satisfy the specified amount. Smaller customers, like retail distributors, do not provide a forecast. As noted in Manary et al. (2019), Intel adopted a process called "Just Say Yes" where Intel strived to satisfy all customer requests. What that meant operationally is that all customer forecasts were treated as requirements to satisfy. Intel's forecasted demand is then a combination of the sum of the customer-provided forecasts shared outside Intel's distribution center replenishment time and its estimate of the non-forecasted sales from smaller customers. Intel judges forecast accuracy by comparing actual customer orders in a week versus the forecasted demand for that week.

Customer Orders represents the actual requests for a given SKU from customers in that given week at that specific distribution center, measured in microprocessor units. As an example, in Table 1 SKU-A-2's week 1 forecasted demand at distribution center ALPHA was 8,949 units while its customer orders were 11,146 units. The forecast demand of 8,949 units was made weeks earlier and the customer orders were placed in week 1. If the DC has sufficient inventory on hand then customer orders equals the unconstrained true customer demand for that week. However, we do not know the DC inventory position so we can only conclude that customer orders are less than or equal to the unknown unconstrained true customer demand.

3   Descriptive Statistics

This section provides context and descriptive statistics to ensure researchers have a common understanding of the underlying data. Manary and Willems (2008), Manary et al. (2009), and Wu et al. (2010) all describe work at Intel to improve supply chain performance by modeling how forecasting and supply chain decisions interact. Therefore, examining forecast error in the data is a natural starting place. Using the familiar mean percentage error (MPE) formula, MPE = $\frac{\text{customer orders} - \text{forecasted demand}}{\text{customer orders}} \times 100\%$, Figure 1 provides a MPE histogram for 97.5% of the records in the data set. The remaining 2.5% comprises the uncharted heavy left tail which was excluded to enable a better visual representation. The data set's maximum overforecast occurs at distribution center DELTA in week 158 when SKU-H-4's forecasted demand of 30,255 units was 1,890 times larger than its customer orders of 16 units. Across all records, the average MPE is -294%. That is, the average forecast error represents a forecasted demand 3.94 times larger than the customer

6

orders. However, when weighted by customer-order volume, the overall weighted average error quotient is lower at -217%, which is still a considerable amount of positive forecast bias to handle.
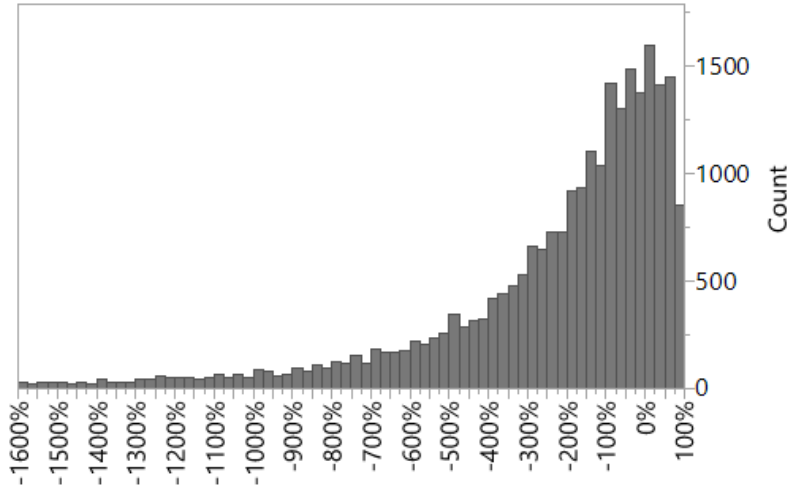

Figure 1: The distribution of MPE scores for 97.5% of the records in the data set. 2.5% of the records have MPE exceeding -1600% but this left tail is excluded to make the figure readable.

Slicing the data by location, Figure 2 displays the average MPE for each distribution center. On average, all distribution centers display a bias to overforecast the eventual customer orders. And although distribution center BETA has a relatively "low" MPE of -48%, it is worth noting the standard deviation at BETA is 181% resulting from a long left (overforecasting) tail.

Table 3 documents a surprising result that SKU level forecasts are not meaningfully correlated. For example, overforecasting a SKU at one location does not meaningfully inform the overforecast or underforecast of that SKU at another location. And while two of the pairs are statistically significant at the p = 0.05 level, denoted by * in Table 3, the magnitudes are not necessarily meaningful.

	ALPHA	BETA	GAMMA	DELTA	EPSILON
ALPHA	1.00				
BETA	0.002	1.00			
GAMMA	0.065*	-0.011	1.00		
DELTA	0.01	0.001	0.003	1.00	
EPSILON	-0.003	0.026	0.066*	0.012	1.00

Table 3: Correlation of MPE error quotients when conditioning at the SKU and week level, where * signifies a p–value < 0.05.
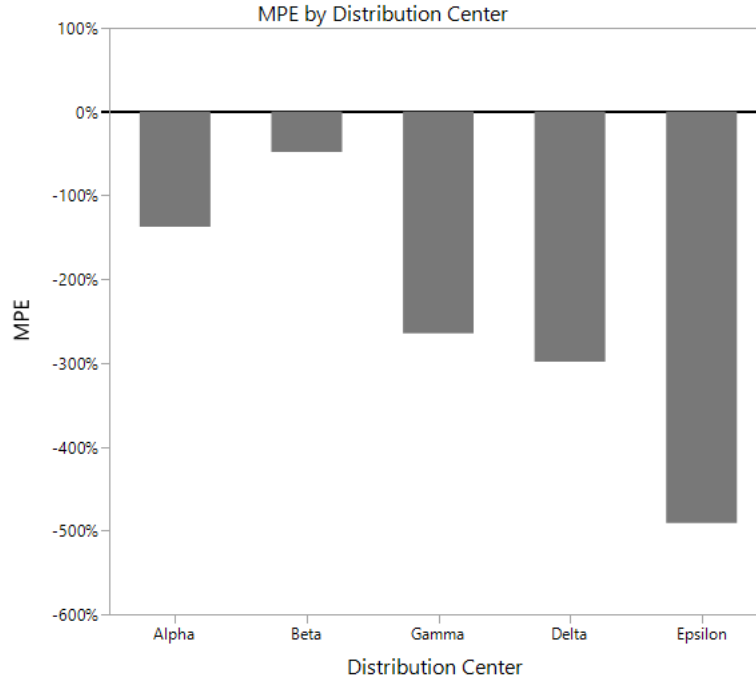
7

Figure 2: Average MPE by distribution center.

4 Intel's Perspective On The Data Set

We present three research areas Intel believes this data set is well suited to address.

4.1 Double booking

Double booking occurs when a customer inflates its forecasts for two SKUs. The simplest case of double booking inflates the forecast for two generations of the same product offering. The data set has the potential to determine whether customers double book during product transitions.

Figure 3 sums customer orders and forecasted demands for each generation of product offering E across the distribution centers, plus a unit difference between aggregate forecasted demands and customer orders. To mitigate the considerable noise in Figure 3, Figure 4 applies a simple spline smoothing transform to the data and the typical product life cycles emerge, along with an upward bump in overforecasting early in the life cycle of generation 3. There is also no discernable trend, with the exception of generation 3, towards more accurate forecasting in aggregate. However, given planning occurs at a SKU-location level, researchers may want to explore the forecasting bias and

8

variance dynamics occurring within the generations at the SKU and location level.

4.2   Price sensitivity

One potential source of forecast bias can arise from not understanding, or miscalling, the price elasticity of customers as new generations of products are released. This data set shows that with each new generation Intel generally adds additional product offerings and expands the highest price points, and even occasionally retires low price points. Real-world performance is measured in many ways, including CPU clock speed, battery life, power consumption, 3D performance, and a host of other software benchmarks. Not all measures improve the same way generation to generation. For example, clock speed may be relatively unchanged one generation to the next but power consumption drops dramatically. Or graphics performance may jump considerably while in the next generation clock speed increases and power consumption is unchanged. In general, it is reasonable to think of each generation as 20% to 40% better than the previous generation.

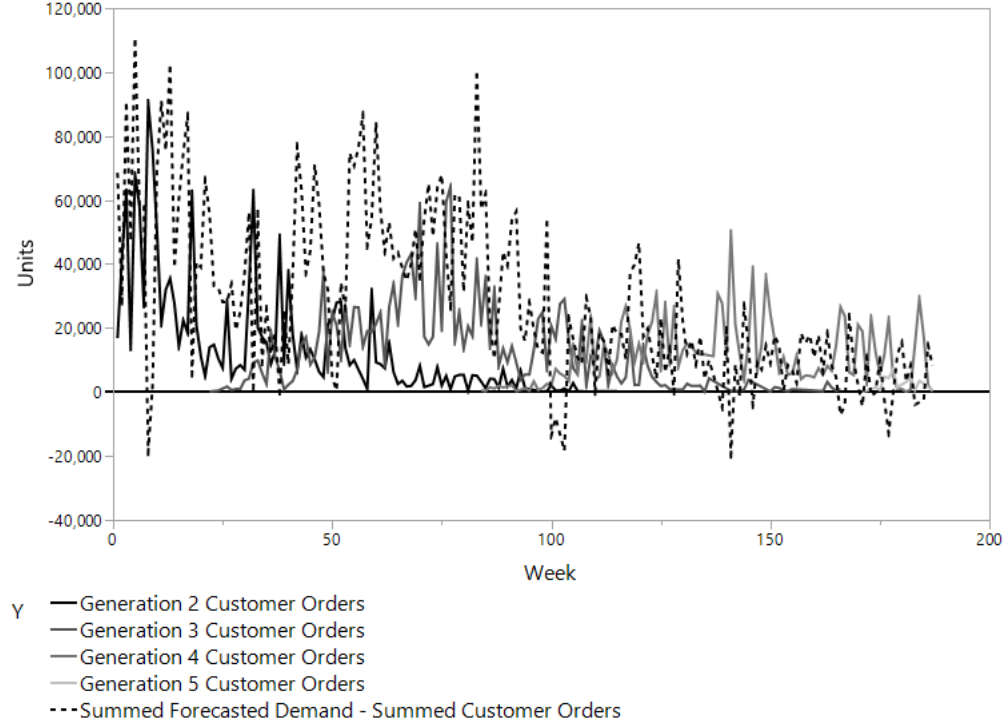The data set's pricing information allows researchers to assess whether customers are more

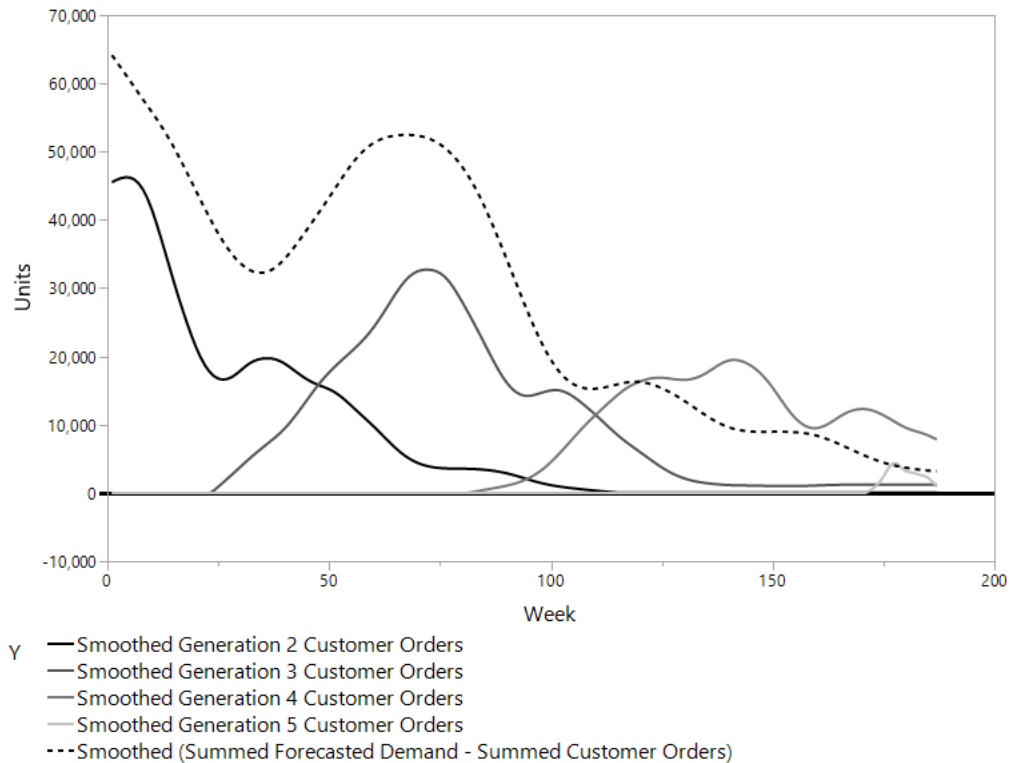Figure 3: Aggregate orders and forecasts for each generation of product offering E.

9

Figure 4: Product offering E's smoothed demand transitions and forecast error.

price elastic or inelastic to performance over time. Another interesting pricing question is whether customers purchase based on relative performance between product offerings in a given generation.

One way to explore this topic could isolate customer behavior across product generations and test for changes in purchasing behavior over time. Intel often sums over all customer orders at a weekly level to identify periods where orders would provide a clean generation-versus-generation view with minimal transition impact. Figure 5 provides a graphic of total customer orders by generation along with a spline-smoothed curve to eliminate the week-to-week noise. Based on Figure 5, the peak of the life cycles occurred around week 71 for generation 3 and week 139 for generation 4. Intel defines a generation's time window as its peak week plus and minus four weeks.

Summing over all other attributes, Figure 6 plots orders by ASP group for the nine-week peak windows of generation 3 and generation 4. In the time between the peak of generation 3 and generation 4 orders, customers shifted their purchases away from ASP groups 1, 7, and 12 and towards ASP groups 4, 8, 10, and 18. Simple observation indicates customers in general moved to higher ASP groups. However, Figure 7 shows that although some of the orders migrated to new
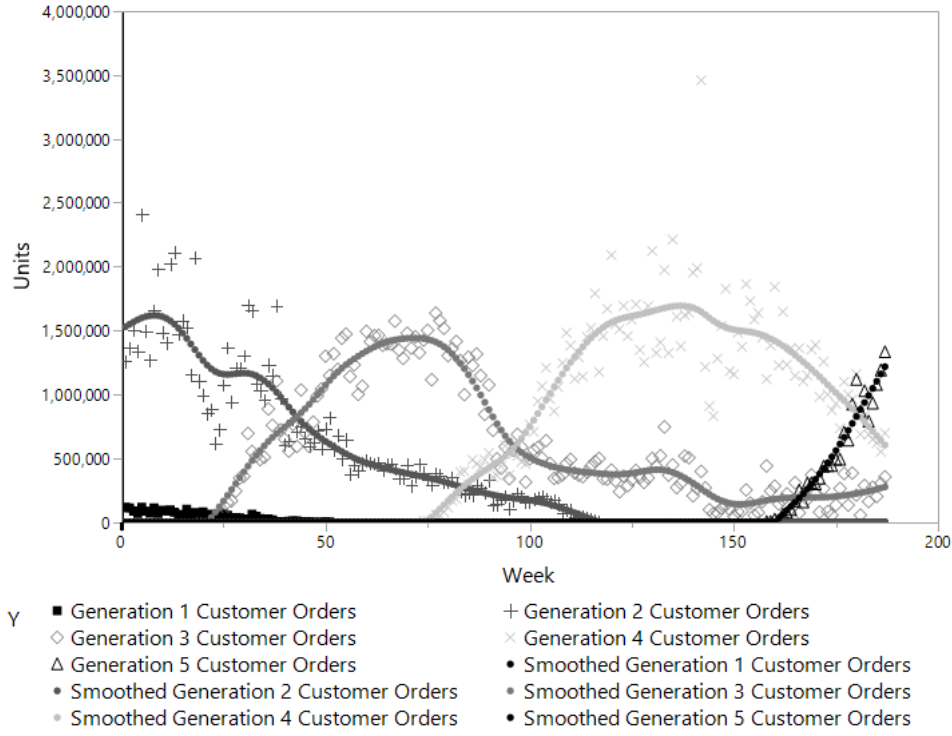
10

Figure 5: Total weekly customer orders and smoothed orders for all generations.

product offerings of D, P, V, and W, there was a significant shift in customer purchases towards the already existing product offerings of C, H, and L.

4.3 Forecasting product mix over time

Intel has to make supply chain decisions at the most granular level in the supply chain. Decisions are made by SKU (product offering and generation) by distribution center by week. A SKU's product mix is its percentage of that week's total units. Correctly forecasting product mix is a challenge given the dynamic nature of available SKUs.

Figure 8 represents the percent of the demand forecasted for each product offering in generation 4's nine-week time window along with the actual percent of customer orders for each product offering. Some surprising behavior comes out of the data. First, the huge shift in demand away from product offering K and towards product offering C seen in Figure 7 was actually forecast well. However, demand forecasts significantly overestimated the mix of demand for product offering F and, on a relative basis, significantly under-called the demand for product offering P. Figure 8
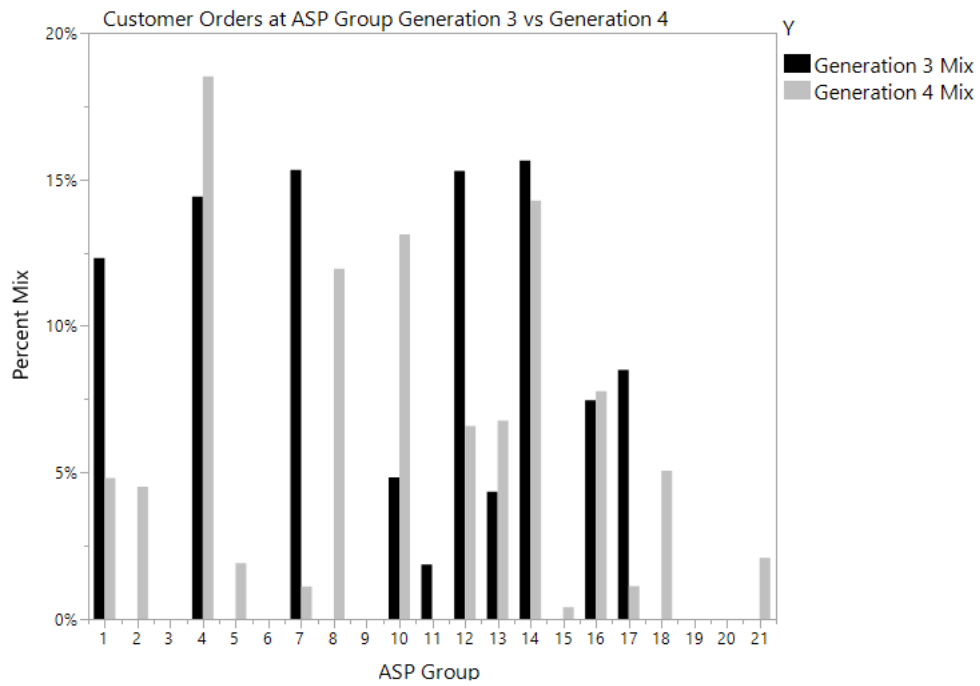
11

Figure 6: Comparing customer orders by ASP group for the generation 3 and 4 time windows.
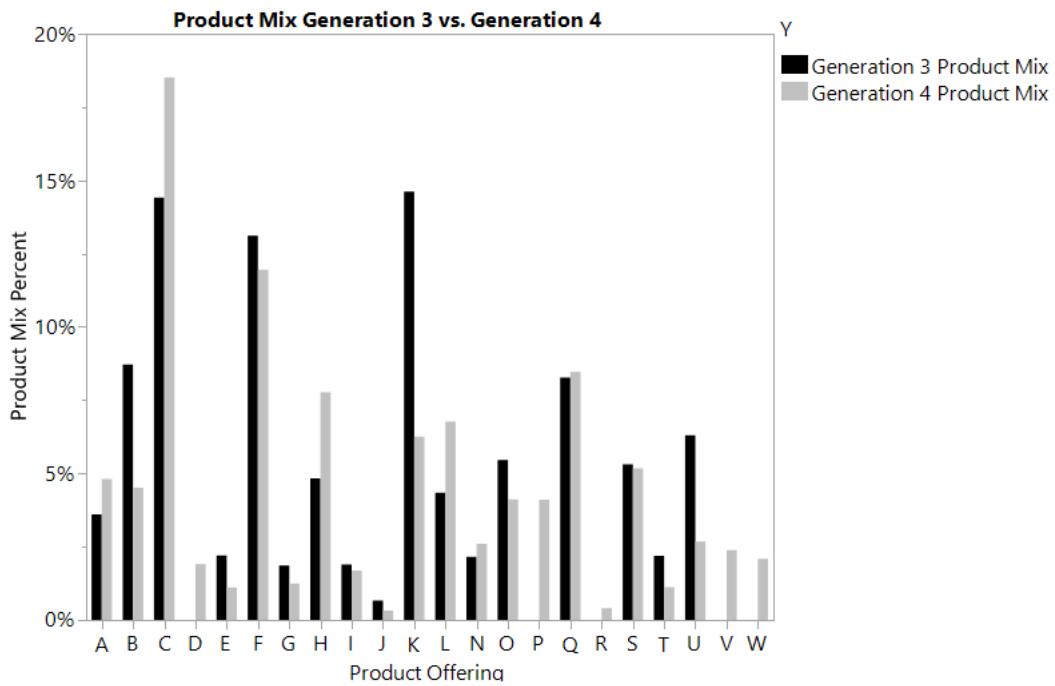

Figure 7: Customer order distribution for the generation 3 and 4 time windows.
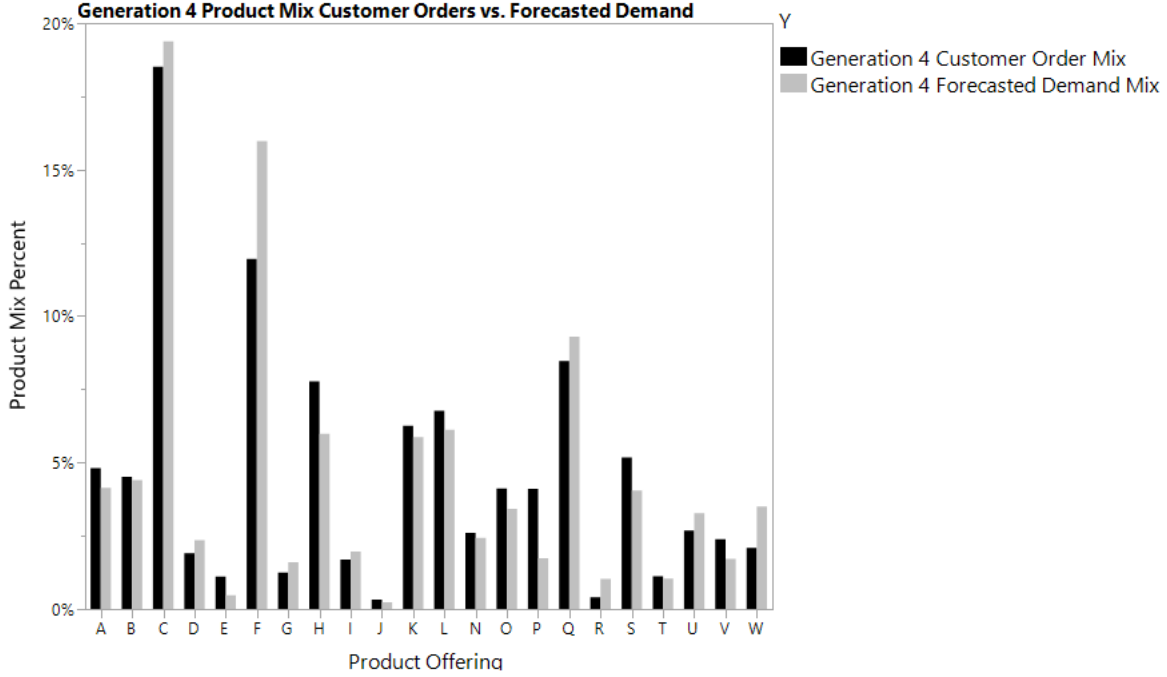
Figure 8: Relative percentage volume for both customer orders and forecasted demand for generation 4's nine-week peak-demand window.

shows that on a relative aggregate basis, even if the magnitude is considerably off, as shown in Figure 1 and Figure 2, it appears Intel's forecast succeeds at estimating which SKUs customers will ultimately select.

5   Conclusion

This data set presents 187 consecutive weeks of Intel microprocessor demand information. There are 86 SKUs in total, and for each SKU the weekly forecasted demand, customer orders, and average selling price category are provided for each of the five distribution centers in the sales geography.

There are many interesting facets of this data set. Intel believes that by providing such an unfiltered view into one geography's data, researchers can test their models and hypotheses in many supply chain domains including forecasting, inventory, pricing, and product assortment. In terms of linking this data set to existing *M&SOM* data set papers, it could be interesting to enrich the semiconductor and computer supply chains in Willems (2008) with the multigeneration setting in this paper. With regards to Acimovic et al. (2019) there is not an exact mapping between SKUs

in the two data sets but it should still be possible to investigate differences between a supplier and a customer operating at different echelons in the same supply chain ecosystem. In the future, a data set that could include specific timing and geographic data would be a valuable addition to the research community.

6   Downloading The Data

The data can be accessed at ⟨INSERT_LINK_HERE⟩. The data is available in two formats. 2019_08_08_CSV_Data.csv is a UTF-8 encoded comma delimited file. 2019_08_08_CSV_Data.xls is a Microsoft Excel 97-2004 workbook.

References

Acimovic, J., F. Erize, K. Hu, D.J. Thomas, J.A. Van Mieghem. 2019. Product life cycle data set: Raw and cleaned data of weekly orders for personal computers. *Manufacturing & Service Operations Management* **21**(1) 171–176.

Manary, M.P., B. Wieland, S.P. Willems, K.G. Kempf. 2019. Analytics makes inventory planning a lights-out activity at Intel Corporation. *Interfaces* **49**(1) 27 pages.

Manary, M.P., S.P. Willems. 2008. Setting safety stock targets at Intel in the presence of forecast bias. *Interfaces* **38**(2) 112–122.

Manary, M.P., S.P. Willems, A.F. Shihata. 2009. Correcting heterogeneous and biased forecast error at Intel for supply chain optimization. *Interfaces* **39**(5) 415–427.

Willems, S.P. 2008. Data set – Real-world multiechelon supply chains used for inventory optimization. *Manufacturing & Service Operations Management* **10**(1) 19–23.

Wu, S.D., K.G. Kempf, M.O. Atan, B. Aytac, S.A. Shirodkar, A. Mishra. 2010. Improving new-product forecasting at Intel Corporation. *Interfaces* **40**(5) 385–396.