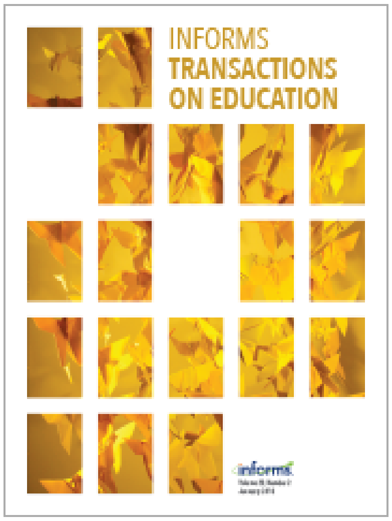


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Case Article—Canyon Bicycles: Judgmental Demand Forecasting in Direct Sales

Christoph Diermann, Arnd Huchzermeier

informs
http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 17, No. 2, January 2017, pp. 58–62
ISSN 1532-0545 (online)

Case Article

Canyon Bicycles: Judgmental Demand Forecasting in Direct Sales

Christoph Diermann,[a] **Arnd Huchzermeier**[a]

[a] WHU—Otto Beisheim School of Management, 56179 Vallendar, Germany

Contact: christoph.diermann@whu.edu (CD); arnd.huchzermeier@whu.edu (AH)

Received: March 4, 2016
Accepted: July 8, 2016
Published Online: January 13, 2017

https://doi.org/10.1287/ited.2016.0165ca

Abstract. We present a comprehensive introduction to judgmental demand forecasting along with a model that allows for effectively debiasing team forecasts and estimating demand distributions. This case is recommended for classes in operations management, marketing, or retail management; two companion papers are ideal for advanced courses (e.g., master's or doctorate programs). We confront students with a demand forecasting problem encountered by Canyon Bicycles, the German premium bicycle manufacturer and online retailer. We present both a model and a process for deriving an accurate judgmental demand forecast. In particular, we demonstrate how one can (i) identify the best team composition, (ii) prepare for and run the forecasting meeting, (iii) debias team forecasts, (iv) estimate demand distributions, (v) deal with heterogeneous product collections, and (vi) judge the quality of forecasted demand distributions.

Supplemental Material: Supplemental material is available at https://doi.org/10.1287/ited.2016.0165ca.

Keywords: judgmental demand forecasting • demand uncertainty • debiasing demand forecasts • forecast accuracy

1. Introduction

In this case study, we describe the development, testing, and successful implementation of a new demand forecast model for the manufacturer and online retailer Canyon Bicycles, a model that led to significant sales growth and reduced supply demand mismatch costs; in particular, aggregated demand forecast accuracy increased from 74% to 97% and company profits increased by 11% (Diermann and Huchzermeier 2016b). (Note that all data have been disguised for the case study.) Canyon Bicycles was founded in 2001 (in Koblenz, Germany) and so can still be viewed as a startup. This company is the acknowledged technology and quality leader in its industry. Besides receiving many design and industry awards, the mountain and road bikes of Canyon Bicycles were used in 2015 by athletes who won the Hawaiian Ironman and races of the European Tour de France stages.

The premium bicycle industry is characterized by high but uncertain growth rates. Sales of Canyon Bicycles have in recent years grown by an average of more than 20% annually. However, not all bikes exhibit the same rate of growth, and a given bike's rate may differ from one season to the next—which means that future demand is poorly predicted by past sales. Intensified competition at the global level has led to shortened product life cycles, expanded product portfolios, and strong technology innovation. In Germany, six major competitors filed for bankruptcy and thereby convinced Canyon Bicycles' management team to make demand forecasting the company's first priority. An additional challenge faced by the firm is that most components require prefinancing and also must be ordered far ahead of the selling season. Capacity and lead-time considerations preclude any in-season orders.

The market dynamics just described severely compromise forecasts that are based directly on past sales, as does the high proportion (more than half) of *new* models in each year's Canyon Bicycles product line. We therefore use so-called judgmental demand forecasts, which are based on the knowledge of company experts who can anticipate the effects (on demand) of strategic changes or market developments.

2. Literature

We follow the preponderance of treatments in the literature and use, as the basis of our demand forecasting model, the arithmetic mean of experts' demand estimates (per bike). However, such mean demand forecasts cannot capture the uncertainty of actual demand. Because direct estimates of uncertainty (i.e., the confidence levels of demand forecasts) are usually too narrow, owing to overconfident team members, demand uncertainty is usually estimated via indirect methods (Soll and Klayman 2004). We have identified three major research streams relevant to our demand forecasting problem—namely, work that addresses (1) estimating mean demand, (2) estimating demand uncertainty, and (3) debiasing team forecasts.

In the first literature stream, mean demand is defined as the arithmetic mean of expert forecasters' demand estimates while demand uncertainty is defined as the standard deviation of those same forecasters' demand estimates at the product level (Fisher et al. 1994). A value of 1.75 has been suggested for the ratio of observed to estimated standard deviations in the forecast errors; thus one could proxy demand uncertainty by (roughly) doubling the forecasters' dispersion (Raman and Hammond 1994, Fisher and Raman 1996). So if we assume that the distribution is normal, then the demand forecast can be defined as the mean demand forecast *plus* the doubled standard deviation of the forecasters' demand estimate. This approach worked well for forecasting fashion items at Sport Obermeyer, for example.

The second research stream extends the first by adding scale—that is, the mean demand forecast—as a control variable to the forecasters' dispersion when predicting demand uncertainty (Gaur et al. 2007). These authors find that scale and dispersion are each positively correlated with demand uncertainty; they also report that the variance of demand increases sublinearly with dispersion yet increases more than linearly with scale. This approach has also been tested using the data set from Sport Obermeyer, where it had the effect of refining parameter estimates.

The third research stream focuses on debiasing the team's mean demand forecast and estimating demand uncertainty based on the previous season's *relative forecast errors* or A/F (actual/forecast) ratios. Cachon and Terwiesch (2013) assume that last season's average relative forecast error—across all products—is similar to its next-season counterpart. To debias the current season's mean demand forecasts, these authors multiply them by the previous season's average A/F ratios. For their proxy of demand uncertainty, Cachon and Terwiesch multiply the standard deviation of the previous season's A/F ratios by the current season's mean demand forecast; with this, they implicitly assume that the randomness of relative forecast errors is related in some way to the randomness of actual demand uncertainty. Retail data for a collection of O'Neill wetsuits is used as an illustration of this forecasting approach.

It turns out that a combination of these models is needed to solve Canyon Bicycles' demand forecasting problem. The Sport Obermeyer case considers neither rapid demand growth nor biases in team members' demand forecasts, and the O'Neill case does not consider new products—for which (by definition) there can be no A/F estimates. We remark that there is great danger in applying the A/F ratio approach blindly, since it can easily lead to the company ordering way too much. Finally, both of these cited cases ignore the question of whether their fitted parameters are stable enough to be used in estimating the next season's demand.

All the claims made in this case study are supported by two companion papers based on three years of observed forecast and sales data for about 100 bikes (Diermann and Huchzermeier 2016a, b). In Diermann and Huchzermeier (2016b), we use the first two years for (in-sample) parameter estimation of our forecast model and the last year for out-of-sample testing. Advanced courses should assign both research papers as additional readings and make cross-references to them during in-class discussion of the case. Thus we encourage discussing, for example, behavioral issues in forecast teams, demand estimation procedures, consideration of old and new bikes, out-of-sample testing, different approaches to estimating demand volatility, and supply chain optimization and coordination; see Diermann and Huchzermeier (2016a, b).

Decision biases also play a key role in demand forecasting models (Schoemaker and Tetlock 2016). As already mentioned, managers are typically overconfident in their estimates of season demand (Fisher and Raman 1996) and even more so with regard to direct estimations of the confidence intervals (Soll and Klayman 2004). It is well documented that managers often suffer from a mean reversion bias when hedging against forecast error (Schweitzer and Cachon 2000, Bolton et al. 2012). The instructor should discuss these persistent biases as well as ways to mitigate them—for example, by utilizing prediction markets (Berg et al. 2008) and other structured forecasting techniques, such as the nominal group technique (van de Ven and Delbeco 1971) and the Delphi method (Dalkey and Helmer 1963). For a comprehensive comparison of various judgmental forecasting methods, see Graefe and Armstrong (2011).

3. The Judgmental Demand Forecasting Approach

Examining the Canyon Bicycles data set reveals that expert forecasters, on average, tend to underestimate actual demand. We therefore debias their mean demand forecast using an approach similar to that described in Cachon and Terwiesch (2013). Because we identified significant differences in the average A/F ratios for certain product segments of Canyon Bicycles, we debias the mean demand forecast—and estimate the demand distribution—using *segment-specific* A/F ratios. In a companion paper, we provide statistical proof that, once the appropriate segmentation is identified, relative average forecast errors are indeed segment specific and remain stable across sales seasons. This finding confirms the hypothesis that using segment-specific relative average forecast errors is an effective way to debias demand forecasts (Diermann and Huchzermeier 2016b).

We estimate demand distributions based on the distribution of the A/F ratios—within each considered segment—while assuming that demand is normally distributed. In the companion paper cited previously, we show that forecast errors of the (segment-specific) debiased demand forecasts do indeed reasonably approximate a normal distribution.

Because one can order only an integer number of bikes, the Poisson distribution should be deployed for very low levels of demand (Cachon and Terwiesch 2017, pp. 451–455). Yet for means greater than 20, the Poisson distribution actually resembles the bell shape of a standard normal distribution. In 2014, only three of the firm's nearly 100 bike models exhibited a demand below 50, and the minimum demand was 29. We therefore decided to use a normal demand distribution for our demand predictions. The data from that case study are disguised but similar to the actual data set, so conclusions drawn from the companion paper are also valid here. When demand is *not* normally distributed, segment-specific empirical distributions can be used instead (as discussed in detail by Schleifer 1992). Schleifer assumes that previous seasons' empirical cumulative density functions are comparable across seasons and therefore uses those seasons' empirical distributions of the A/F ratios to predict the current season's demand distributions. To check the accuracy of the empirical distribution function, we suggest that the targeted and obtained service levels be compared at the end of the season—thereby testing the validity of the assumption that empirical distribution functions are comparable across seasons.

In this case study, we expose students to analytical tools that can be applied to three important tasks: identifying the best segmentation of a product line; estimating demand uncertainty; and debiasing the mean demand forecast. We also provide simple tests that help students develop intuitive ways of judging the quality of estimated demand distributions.

4. Teaching Plan

The case study can be used in an advanced operations management, marketing, or retail management class that discusses (judgmental) demand forecasting when actual demand is observable (note that Canyon Bicycles sells almost all of its products online and tracks lost sales for each bike). As we stated previously, the companion papers can be used in more analytical courses. Related case studies are L.L.Bean (Schleifer 1992), Sport Obermeyer (Raman and Hammond 1994), and O'Neill (Cachon and Terwiesch 2013). The L.L.Bean case study introduces students to the concept of debiasing demand forecasts using different A/F ratios for new and old products. The Sport Obermeyer case discusses team-based forecasts and the issue of overconfidence when estimating demand volatility. The O'Neill case presents a scale-based estimator of demand volatility as well as an estimator based on a *single* A/F ratio for all components of a product collection.

We build on these previous cases by introducing two new concepts: segment-specific A/F ratios, which discourage unduly amplified aggregate forecasts and the attendant overstocking or understocking of products; and the use of actual sales data to test the derived demand distributions. In our case, debiasing products using A/F ratios for old and new products did not lead to accurate demand forecasts. Our demonstration that some product segmentations lead to robust demand estimates across seasons is deferred to a companion paper; see Diermann and Huchzermeier (2016b).

In this case study we describe in detail the previous, CEO-driven consensus demand forecast as well as the new, A/F ratio-based demand forecasting process now employed by Canyon Bicycles. Students receive team-based forecasts for eight bikes that can be segmented at three levels, after which they must use actual season sales data to validate their choice of segmentation (out-of-sample-testing). Students are also asked to benchmark the "old" consensus forecast, as well as their "new" (segment-specific) debiased forecasts, against a simple top-down forecast derived from historical data on relative product shares and other heuristics.

5. Classroom Experience

One of the main reasons for developing this case study on judgmental demand forecasting for direct sales is that many of our bachelor's and master's students are engaged in developing an online startup company during their studies. Famous examples are, among others, Zalando (Europe's largest and fastest-growing online fashion retailer), Rocket Internet (a venture builder), Audibene, known in the United States as hear.com (an online retailer of hearing aids), HelloFresh (an online food delivery service), SumUp (an online payment service provider), and Amorelie (an online retailer of adult products). Hence our students naturally show a strong interest in research related to the management of such online businesses and, in particular, to topics dealing with process management and improvement. This case serves as a real "eye opener" for many students, especially when they return home after class to continue tinkering with plans for their own startup company. When the instructor is able to position the case in this way, it is then great fun and relatively easy to teach students.

Our case study has been tested with 79 M.Sc. candidates—a large fraction of our annual student intake, and all of whom already have a bachelor's degree in management or economics—in an operations management course. During the class session, the head of production and logistics for Canyon Bicycles was also present. All students were familiar with basic statistical concepts and with the newsvendor model. The opening scenario, in which hundreds of millions of euros must be committed to orders for mostly new bikes, poses a real challenge. Only after students benchmark the various forecasting approaches do they come to understand the value of product segmentation and of judgmental demand forecasting that employs company experts. A key insight is that a careless or even casual attitude about the aggregate number of bikes ordered can have severe commercial and economic consequences for this family owned company.

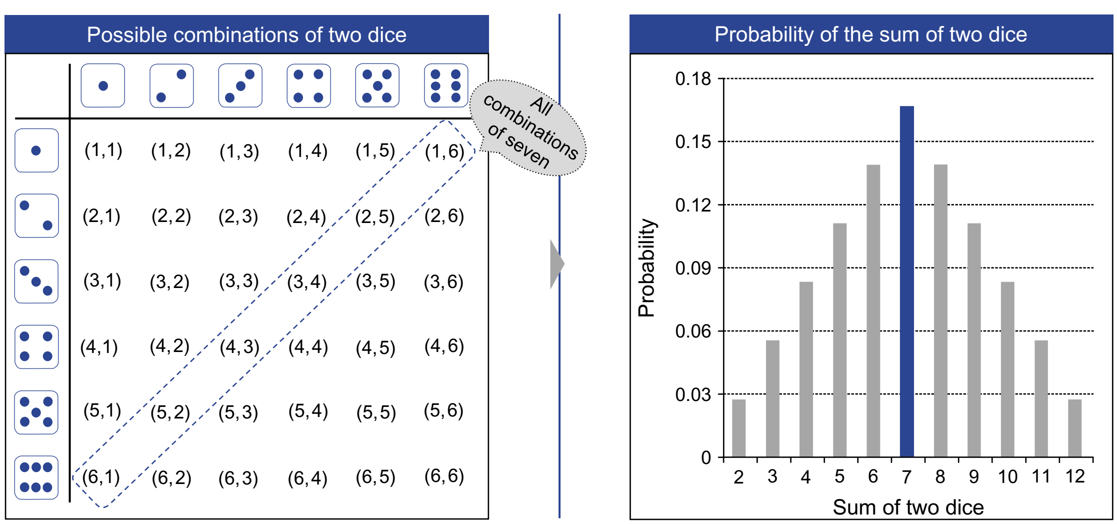
To help students better grasp the idea of how to test the accuracy of derived demand distributions, we discuss how they could quickly spot a rigged dice game—for example, "Craps" played in a casino: the most frequent score should be 7, and there should be a triangular distribution with a low of 2 and a high of 12, see Figure 1 and teaching note for more details (available as supplemental material at https://doi.org/10.1287/ited.2016.0165ca). Any deviation from this frequency distribution would indicate that the dice are rigged. Once they know the distribution, students are usually quick to understand this statistical reasoning and confirm that they consequently face a low risk of being cheated.

Similarly, the actual demand for Canyon Bicycles is expected to be symmetrically distributed around the mean demand forecasts—that is, the demand for half of all bikes should be lower than their mean forecast. This would indicate that the predicted means are quite accurate. To test the accuracy of the entire demand distribution, we compare the target and obtained service level of different percentiles. It is only by deriving accurate demand distributions that Canyon Bicycles can maximize its supply channel profit and minimize its risk exposure. As in the game of Craps, betting on the "wrong" number (here, the ordered components) leads to severe financial consequences, especially when the underlying distribution is intentionally shifted or poorly predicted.

We have observed that students are skeptical when confronted with the initial task of applying judgmental demand forecasting in this particular setting. Rapid demand growth, changing product assortments, and biases in team members' demand forecasts create enough complexity that students long for a "guiding method." That is why the case is written in the style of a tutorial; references to inappropriate heuristics are made only in side comments, so as not to distract students from the main story. After the case discussion and analysis (and a debriefing session), most students appreciate this case for two main reasons. First, its highly adaptable model structure allows one to incorporate the key features of many demand forecasting problems. Second, it has broad applicability in practice (e.g., for startup companies selling online).

In most cases, we use one 90-minute class session to teach the case study; this involves discussing the initial setting (which includes a visit to the company's website), describing both the old and the new forecasting approaches, and testing demand forecast accuracy. The debriefing slides are used in the session wrap up, which lasts from 15 to 20 minutes. Our main purpose there is to walk students through the profit calculations based on actual sales data for the various forecasting approaches. These calculations are straightforward and are repeated several times, although most students are already able to derive the correct profit estimates. The top students are generally better at identifying which segmentation strategy best matches the normality assumption. A tutorial could well use our debriefing slides as standard lecture notes; in that event, however, students should carefully read the case in advance so as to maximize their learning experience.

Figure 1. Probability Distribution of Two Fair Dice



Students submit their two-page group case assignments (with any number of exhibits) at the start of the course. A standard grading sheet—based on the assigned questions given in the teaching note—is provided as feedback to each team. We believe that students should discuss all aspects of the forecasting process and not simply focus on the profit calculations.

6. Concluding Remarks

The case provides a useful contrast to at least three other sections of our operations management course, in which we discuss capacity options in capital-intensive industries without overbooking (e.g., at CargoLifter; Spinler and Huchzermeier 2006) and with overbooking (e.g., at Lufthansa Cargo; Hellermann et al. 2013). That course's third section is devoted to promotional demand forecasts; at the German retailer Metro, for example, a stockpiling model is used to schedule promotional sales optimally by tracking the purchase behavior of price-sensitive customers (Breiter and Huchzermeier 2015). These different cases and papers help students realize that, in practice, there is a wide range of important forecasting problems that can often lead to conflicts between supply chain partners and for which there exist different viable solution approaches. Some approaches rely on deep insight into customer behavior to explain observed demand volatility, although such connections can rarely be made with much confidence. It is almost always a more practical (and reliable) course of action to rely on company experts when predicting future demand and before committing to orders (or to capacity investments). That approach typically leaves students with the uneasy feeling of "just having to live with the consequences" of such irreversible decisions. Yet the main takeaway from the Canyon Bicycles case is that, to beat the competition, one had "better make good decisions."

References

Berg JE, Nelson FD, Rietz TA (2008) Prediction market accuracy in the long run. *Internat. J. Forecasting* 24(2):285–300.

Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and students as newsvendors. *Management Sci.* 58(12):2225–2233.

Breiter A, Huchzermeier A (2015) Promotion planning and supply chain contracting in a high-low pricing environment. *Production Oper. Management* 24(2):219–236.

Cachon G, Terwiesch C (2013) *Matching Supply with Demand: An Introduction to Operations Management*, 3rd ed. (McGraw-Hill/Irwin, Boston).

Cachon G, Terwiesch C (2017) *Operations Management*, 1st ed. (McGraw-Hill, New York).

Dalkey N, Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Management Sci.* 9(3):458–467.

Diermann C, Huchzermeier A (2016a) Estimating demand uncertainty using dispersion of team forecasts or distributions of forecast errors. Working paper, SSRN: http://ssrn.com/abstract=2782402.

Diermann C, Huchzermeier A (2016b) Judgmental demand forecasting for online sales of collections: Segmentation by product type or age? Working paper, SSRN: http://ssrn.com/abstract=2782002.

Fisher M, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* 44(1):87–99.

Fisher M, Hammond J, Obermeyer W, Raman A (1994) Making supply meet demand in an uncertain world. *Harvard Bus. Rev.* 72(3):83–93.

Gaur V, Kesavan S, Raman A, Fisher ML (2007) Estimating demand uncertainty using judgmental forecasts. *Manufacturing Service Oper. Management* 9(4):480–491.

Graefe A, Armstrong JS (2011) Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *Internat. J. Forecasting* 27(1):183–195.

Hellermann R, Huchzermeier A, Spinler S (2013) Options contracts with Overbooking in the Air Cargo Industry. *Decision Sci.* 44(2):297–327.

Raman A, Hammond JH (1994) Sport Obermeyer Ltd. Harvard Business School Case 695-022 (Revised August 2006).

Schleifer A Jr (1992) L.L. Bean, Inc.: Item forecasting and inventory management. Harvard Business School Case 893-003 (Revised September 1993)..

Schoemaker P, Tetlock P (2016) Superforecasting: How to upgrade your company's judgment. *Harvard Bus. Rev.* 94(5):72–78.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Soll JB, Klayman J (2004) Overconfidence in interval estimates. *J. Experiment. Psych.: Learn., Memory, and Cognition* 30(2): 299–314.

Spinler S, Huchzermeier A (2006) The valuation of options on capacity with cost and demand uncertainty. *Eur. J. Oper. Res.* 171(3): 915–934.

van de Ven A, Delbeco AL (1971) Nominal versus interacting group processes for committee decision-making effectiveness. *Acad. Management J.* 14(2):203–212.