Introduction. The newsvendor model is a cornerstone of the Operations Management discipline. A four-digit number of academic papers extends it, and most major business schools teach the model as a part of their core MBA curriculum. For retailers that face short product lifecycles, long lead-times and high product variety, the model is more relevant than ever. Given this, one would expect the model to be frequently applied in practice, with its parameters estimated and the optimal quantity ordered. Yet the reports of such applications are scarce. Arguably, this is because a key input for the newsvendor model, the demand distribution, is difficult to obtain in practice. In particular, for new products, we are not aware of widely adopted methods for demand distribution forecasting.

A promising way to forecast demand for new products is based on “the wisdom of crowds”: collecting and aggregating quantitative point forecasts of demand. Giving such inputs, however, might still be too complicated for many potential forecasters, excluding valuable opinions. For example, product designers may not be exposed regularly to sales figures, thus not being comfortable in expressing their estimate of product attractiveness in form of a precise quantity; yet their product insight might prove beneficial. To include the opinions from such forecasters, we propose to go from quantities to less detailed inputs.

When the company regularly introduces new products in established categories, the case we’ll be focusing on, a natural way to simplify the inputs is to ask each forecaster to rank the products within a category. Such subjective rankings, however, cannot be used directly as the sole input for a demand distribution forecasting procedure, so a method that maps rankings to quantities is needed. To come up with such a method, we decompose the demand vector into three parts: aggregate demand for the category, ordered proportions and the ranking. We use the historical data for the category to forecast the first two components and subjective rankings for the last one.
Contribution. Firstly, we establish a modular framework for demand distribution forecasting that allows to use a new kind of inputs: subjective rankings. Secondly, we devise multiple forecasting specifications for the components of the framework, in particular, for ordered proportions and rankings. Thirdly, we propose a new metric for evaluation of distribution over rankings: a Spearman-Brier score.

Forecasting framework. For tractability, we make an assumption that the aggregate demand, ordered proportion and ranking distributions are independent. This suggests a simple simulation-based algorithm: sample an ordered proportion vector, an aggregate demand value and a ranking, reorder the proportions according to a ranking and multiply by aggregate. To complete the algorithm, one needs to specify the distributions of the components. For each of the three, we propose multiple specifications; these can be plugged into the algorithm independently of each other.

An important practical issue in deploying the framework is that the number of products might be different from season to season. The specifications for aggregate and ordered proportion components need to reflect that. For aggregate demand, we consider two specifications: one is a log-log regression of aggregate demand on the number of products, another is based on a logistic curve, with the expected aggregate demand being a convex-concave function of the number of products. For ordered proportion, we propose four candidates: ordered exchangeable Dirichlet distribution, exchangeable Dirichlet in barycentric coordinates, log-ratio multinomial logistic model and exponomial choice model.

For the ranking component, we propose two candidate specifications: one is the empirical distribution over all subjective forecast, another is the Plackett-Luce model.

Evaluation. We evaluate the out-of-sample performance of the resulting distribution forecasts for all the three parts of the decomposition and as a whole. We consider various metrics in doing so. One family of metrics is given by strictly proper scoring rules. For quantitative components, these are mean squared error and continuous ranked probability score. For the ranking component, we propose a new proper scoring rule called Spearman-Brier score that has the same scale as a regular Brier score: 0 for the perfect forecast, 1 for the worst and 0.25 for the uniform over all rankings.

Another set of metrics is based on point and interval forecasts. These include, for quantitative outputs,
the following: mean absolute percentage error, predictive interval coverage and newsvendor profit (for the complete forecast). For rankings, the proportion of top-1 guesses and the Spearman correlation of the actual ranking with a Borda aggregate of subjective rankings is used.

**Empirical illustration.** We apply the framework using the historical sales data and subjective ranking inputs collected at a mid-sized European retailer. We are considering the season that entails the period from January to June 2017. The forecasting session took place in the end of August 2016. There are 55 new products in total, grouped into 12 categories. The 52 forecasters ranked the products within each category, giving 2754 inputs in total. To each of the 12 categories, historical demand data for the past season on similar products, according to company’s internal classification, are linked. The demand is computed over the period from January to June 2016. In total, there are 115 products in the historical demand data set.

We are comparing three scenarios: “no ranking”, in which that all rankings are equally likely, “known ranking”, in which the true actual ranking is assumed to be known in advance, and “estimated ranking”, meaning that the ranking is estimated from the subjective inputs. Comparing estimated ranking scenario to the known ranking one allows to assess the potential benefit of improving the quality of ranking estimation; comparing estimated ranking to no ranking allows to estimate the benefit of using ranking inputs in the first place.

For a particular specification of the components, we observe the following preliminary findings: the aggregate demand can be predicted with a MAPE of 34%, the proportion – with a MAPE of 73% and the ranking with a Spearman-Brier score of 0.23. In the “no ranking” scenario, the MAPE is equal to 985%, which goes down to 608% for the “estimated ranking” and to 102% for the “known ranking” one.