# COMBINING POPULATION STATISTICS, CONJOINT DATA AND PURCHASE HISTORY IN PRICE OPTIMIZATION

**L. Valkonen[1], S. Tikka[1], J. Helske[1] and J. Karvanen[1]**

[1] Department of Mathematics and Statistics
University of Jyvaskyla, Finland
e-mail: juha.t.karvanen@jyu.fi

## Abstract

With multiple data sources, it is possible to address estimation problems that would otherwise remain unsolvable. The inherent differences between populations or the sampling methods provide a major challenge for data-fusion. We demonstrate how causal modeling and Bayesian estimation can overcome these challenges in a business scenario.

The pricing of products and services is one of the most important decisions that almost every company faces. Our scenario is motivated by a real business case and focuses on the pricing of a subscription-based service, such as video or music streaming, an audiobook service, or a digital newspaper. The subscription is automatically renewed every month unless canceled by the consumer. In price optimization (Phillips, 2021), a company has to estimate the impact of a price change on the behavior of both current customers and potential new customers. Thus, price optimization is essentially a causal problem.

The data sources for the price optimization scenario are the following:

**Population statistics** are available from Statistics Finland and provide information on the joint distribution of age, gender, and location in the target population.

**Conjoint data** originates from a study where the price is varied and a group of customers is asked to make imaginary purchases in an artificial setup (Rao, 2014).

**Purchase history** contains customer-level data on the subscription periods and is collected by the company as a part of daily operations.

The proposed approach (Valkonen et al. 2023) consists of four steps: 1) The causal relations of the purchase process are described in the form of a directed acyclic graph (DAG). 2) The causal effect of the price on purchases is identified from the data sources presented in a symbolic form (Tikka et al., 2021). 3) A hierarchical Bayesian model is fitted to estimate the causal effect based on the obtained identifying functionals. 4) The posterior distribution of the optimal price is found by maximizing the expected gross profit defined as a function of the price and the purchase probabilities estimated in step 3. The approach is demonstrated with simulated data resembling the features of real-world data.

**Keywords:** Bayesian model, Causal inference, Data-fusion, Demand estimation, Transportability

## References


Phillips R. L. (2021) *Pricing and Revenue Optimization*. Stanford University Press.

Rao V. R. (2014) *Applied Conjoint Analysis*. Springer.

Tikka S., Hyttinen A., Karvanen J. (2021) Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software*, 99(5):140.

Valkonen L., Tikka S., Helske J., Karvanen J. (2023) Price optimization combining conjoint data and purchase history: A causal modeling approach, *arXiv:2303.16660*, https://arxiv.org/abs/2303.16660.