Sign and order constraints in hierarchical prior distributions and its benefits for counterfactual predictions

Max J. Pachali[1], Marco J.W. Kotschedoff[1], Peter Kurz[2] & Thomas Otter[1]

[1]Goethe University Frankfurt, [2]bms marketing research + strategy

May 30, 2019

Hierarchical Bayesian models and market level inference

- Applications of hierarchical models have become very popular in marketing, e.g. based on household scanner panel data or from discrete choice experiments
- Derive optimal prices or product configurations representative for population of consumers
- Estimate of consumer preference distribution will be largely affected by specification of hierarchical prior in typical applications
  - $\rightarrow$ small individual information relative to size of model
  - $\rightarrow$ limited price variation in observational data

Standard hierarchical priors and economic rationality

- Unfortunately, standard hierarchical prior distributions often lack economic rationality:
  - $\rightarrow$ Half of consumers in the car market dislike fuel economy in BLP's (1995) estimated distribution of marginal utility
  - $\rightarrow$ Dube, Hitsch and Rossi (2010) find posterior support for positive price coefficients in the inferred heterogeneity distribution
- Makes it difficult, if not impossible, to derive counterfactual predictions from the model
  - $\rightarrow$ Posterior violations of strictly ordered preferences may result in optimal prices that lack face validity (e.g. "cars with less fuel economy being more expensive")
  - $\rightarrow$ Posterior support of positive price coefficients imply infinite prices
- Solution?

Economically faithful hierarchical priors

- Sign and order constraints dogmatically express prior knowledge about the support of a distribution
- E.g., that the price parameter in an indirect utility function is negative or that a consumer prefers a more fuel efficient to a less fuel efficient car for sure, everything else equal
- Prior constraints avoid the extrapolation of parametric assumptions into directions that violate theoretical knowledge
- The goal is a hierarchical prior that is:
    1. maximally flexible regarding some aspects of the population distribution of preferences (e.g., mixture of normals vs. simple normal functional form) **and**
    2. heavily constrained by economic theory regarding other aspects of this distribution

Situations that call for prior constraints

- Suppose a multi-product monopolist offering low and high quality tier products with market prices p_l, p_h and $p_l << p_h$ ("vertical differentiation")
- Even if products are temporally on price promotions, not many consumers can afford the higher quality as $p_l << p_h$ remains
- Difficult (if not impossible) to disentangle consumer's preferences for quality from budget restrictions or price sensitivity
- Prior constraints function as identification restrictions in such cases and rule out likelihood explanations that violate economic principles (e.g. "budget constrained consumers prefer the low quality product")

Goals

1. Illustrate the benefits of economically motivated sign and order constraints for:
   - structural parameter estimates (e.g. brand preferences and price sensitivity) and
   - counterfactual predictions (e.g. marginal cost estimates)

   using an illustrative case study on households' purchases of fresh hen's eggs in Germany
2. Briefly discuss a methodological framework that efficiently implements prior constraints in hierarchical Bayes models as proposed in Pachali, Kurz and Otter (2018)

The "Eggs-Paper": Kotschedoff and Pachali (2019)

Forthcoming, *Marketing Science, Special Issue on Consumer Protection*

- Analyze the impact on consumer welfare of the EU-wide ban on cages for egg-laying hens based on purchase data of German households
- In Germany, eggs are clearly labeled and vertically ordered with regard to perceived animal welfare
- In equilibrium, low income households are especially worse off while higher income households tend to benefit due to falling prices of higher quality products (regressive effect)
- Tailored subsidy scheme to soften the regressive effect
- Less competition in retail market structure and even higher minimum quality standards amplify regressive effect

Ethical production standards

vs.

Regulation

vs.

Consumer welfare implications?

German egg market

- Eggs are differentiated in animal welfare:

Egg label	Hens per $m^2 (\approx 11 ft^2)$	Surface per hen in cm^2	Outdoor area per hen in m^2	Additional points
Organic	6	1667	4	Organicly fed, no regular use of antibiotics
Free-range	9	1100	4	Live in open barns
Barn	9	1100	0	Live in open barns
Battery	18	550	0	Live in cages

Source: `http://www.deutsche-eier.info/die-henne/haltungsformen/`; accessed 2 March 2016

- Consumers associate four breeding categories with different quality levels:

$$Battery\ eggs \precsim Barn\ eggs \precsim Free\text{-}range\ eggs \precsim Organic\ eggs \quad (1)$$

Regulation:

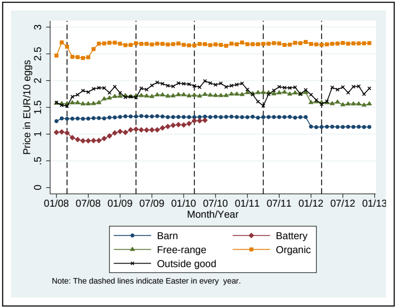
- EU-wide ban of battery eggs since 2012
- In Germany already since 2010

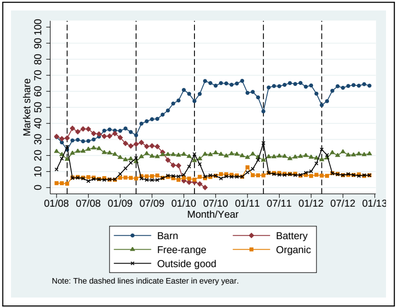
Nielsen Homescan data: 2008-2012

Homescan consumer panel:

- Consumer demographics: Income, age, household size etc.
- Purchase and product information: Egg label, price, store, package size etc.
- Outside option: Aggregate of purchase incidents in related product categories at market-share weighted prices (boiled and painted eggs as well as eggs from other type of poultry, e.g. quails)
- Purchases at top ten retailers (about 75% of all egg purchases)
- 6,961 households (# egg purchases ≥ 4)
- 380,790 observations

Monthly market shares and average prices across egg categories





No overlap of prices between higher quality tiers

$\rightarrow$ Identification?

Empirical demand framework

- Household i's indirect utility from egg product g in retail chain l at period t is:

$$U_{iglt} = \begin{cases} \gamma_{i,g}\mathbf{1}\{t = regular\} + \tilde{\gamma}_{i,g}\mathbf{1}\{t = Easter\} + \\ + \bar{\tilde{\gamma}}_{i,g}\mathbf{1}\{t = Christmas\} + \alpha_i p_{glt} + \beta_i \mathbf{1}\{units_g = 6\} + \\ + \psi_{i,l} + \varepsilon_{iglt}, \quad \text{if } g = Battery \\ \gamma_{i,g}\mathbf{1}\{t = regular, t < RC\} + \gamma^{RC}_{i,g}\mathbf{1}\{t = regular, t \geq RC\} + \\ + \tilde{\gamma}_{i,g}\mathbf{1}\{t = Easter\} + \bar{\tilde{\gamma}}_{i,g}\mathbf{1}\{t = Christmas\} + \\ + \alpha_i p_{glt} + \beta_i \mathbf{1}\{units_g = 6\} + \psi_{i,l} + \varepsilon_{iglt}, \quad \text{else} \end{cases} \tag{2}$$

- $g \in \{Battery, Barn, Free\text{-}range, Organic\}$, $l \in \{1, \ldots, 10\}$
- $\psi_{i,l}$ denotes household i's preference parameter for retail chain l

$\rightarrow$ Identification constraints?

What do we know about indirect utility a priori

- Price coefficient: $\alpha_i \leq 0$
- Package size six coefficient: $\beta_i \leq 0$
- Egg label intercepts:

$$\begin{aligned}
& \gamma_{i,Battery}, \tilde{\gamma}_{i,Battery}, \bar{\tilde{\gamma}}_{i,Battery} \gtrless 0 \\
\text{Regular: } & \gamma_{i,Battery} \leq \gamma_{i,Barn} \leq \gamma_{i,Free\text{-}range} \leq \gamma_{i,Organic} \\
\text{Easter: } & \tilde{\gamma}_{i,Battery} \leq \tilde{\gamma}_{i,Barn} \leq \tilde{\gamma}_{i,Free\text{-}range} \leq \tilde{\gamma}_{i,Organic} \\
\text{Christmas: } & \bar{\tilde{\gamma}}_{i,Battery} \leq \bar{\tilde{\gamma}}_{i,Barn} \leq \bar{\tilde{\gamma}}_{i,Free\text{-}range} \leq \bar{\tilde{\gamma}}_{i,Organic} \\
\text{Regime change: } & \gamma_{i,Battery} \leq \gamma^{RC}_{i,Barn} \leq \gamma^{RC}_{i,Free\text{-}range} \leq \gamma^{RC}_{i,Organic}
\end{aligned}$$

- Retail chain coefficients: $\psi_{i,l} \gtrless 0$ for all l

Estimation technique:

- Hierarchical Bayesian multinomial logit model with a mixture of normals first-stage prior as in Pachali et al. (2018) and Rossi et al. (2005)

What happens for **one household**?

- Let's first look into individual level posterior estimates of a household that mainly purchased lower quality eggs
- Contributed 124 observations

	Barn, 10	Battery, 10	Outside, 10	Barn, 6	Free-range, 10
# Purchases	104	10	8	1	1

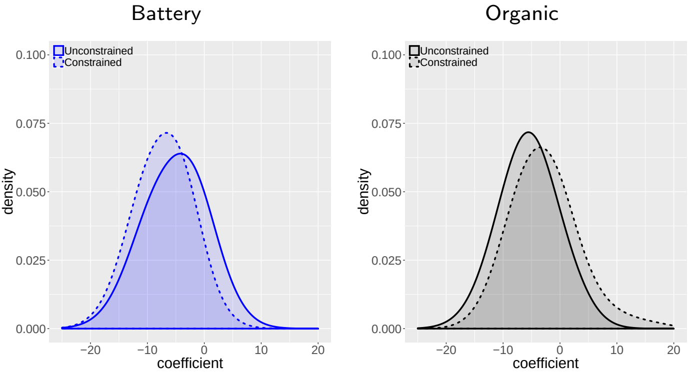
Table: Number of purchases of egg products observed for "household 6"

Comparing posterior estimates for **one household** b/w unconstrained & constrained models

Quantiles	Unconstrained			Constrained		
	Price	Battery	Organic	Price	Battery	Organic
5%	-0.4	-0.9	-11.9	-4.6	-5.1	-3.1
25%	0.5	0.1	-9.7	-4.0	-4.3	-2.5
50%	1.1	0.9	-8.2	-3.6	-3.7	-2.0
75%	1.8	1.7	-6.7	-3.3	-3.0	-1.5
95%	2.7	2.8	-5.0	-2.8	-2.1	-0.8
Mean	1.1	0.9	-8.3	-3.6	-3.6	-2.0
Stand. Dev.	0.9	1.1	2.2	0.5	0.9	0.7

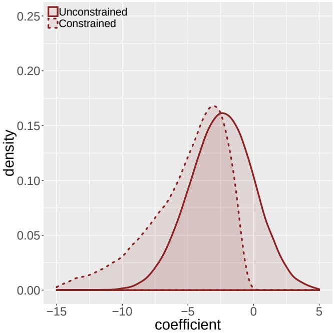
- The unconstrained model estimates large valuations of poor quality tiers as well as low valuations for higher quality tiers
- The price coefficient is unidentified for this household and positive on average in the unconstrained model
- Estimates are economically more convincing in the constrained model!

Market perspective: consumer preferences for egg labels b/w unconstrained & constrained models



→ We overestimate taste for low quality eggs and understimate taste for high quality eggs in the unconstrained model

Market perspective: consumer distribution of price coefficients b/w unconstrained & constrained models



$\rightarrow$ Price sensitivity is not identified in the unconstrained model, predicting ca. 20% of consumers with positive price coefficients

Implications for counterfactual predictions: Marginal cost estimates

- Retailers' marginal costs for egg products are unobserved by researchers
- Given observed retail egg prices in 2008, however, we can estimate marginal costs using a model of retail competition:

$$s(p) + [\Omega * \Delta](p - c) = 0, \tag{3}$$

  where $s(p)$ is the vector of market share; Ω is the product ownership matrix and Δ is the partial derivative matrix of market shares w.r.t. price
- $s(p)$ and the partial derivative matrix Δ are a function of the consumer preference distribution that can be derived from the **unconstrained** or **constrained** version of the model

Comparing marginal cost predictions

	Unconstrained	Constrained
Battery 10 units	0.02	0.48
Barn 10 units	0.18	0.73
Free-range 10 units	-0.25	0.90
Organic 10 units	-0.47	1.45

Table: Market-share weighetd average marginal cost estimates (across egg products offered by the 10 retail chains)

$\rightarrow$ Constrained model implies marginal costs that are much more in line with prior expectations, common sense

Predictive model fit can be really misleading

- Log marginal likelihood based on Newton-Raftery:

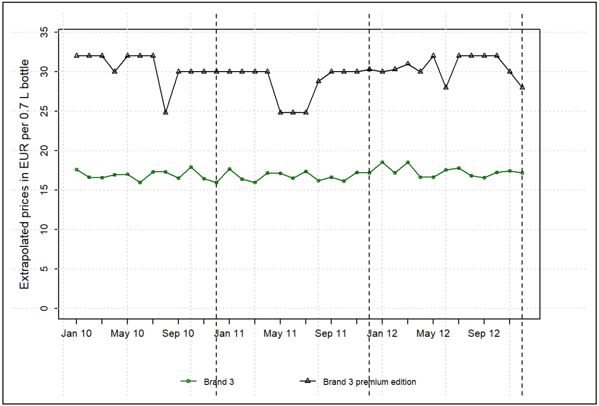
	Value
Unconstrained one normal component	-94979.97
Constrained one normal component	-98245.03

- Often models that defy common sense appear to fit the data better than theory-constrained models: There is no likelihood punishment for explaining the data in an economically misleading way

Another example: the German whisky market

- The German whisky market is dominated by seven major national brands: Jim Beam, Jack Daniel's, Ballentine's , . . .
- Each brand offers a quality-differentiated product line to segment consumers with respect to their willingness-to-pay for quality
- Market prices indicate the strong vertical differentiation between quality tiers (base vs. premium) of the same brand

Prices of baseline vs. premium quality



- Despite temporal price drops, not many consumers switch to the premium product in the data

Estimating premium coefficients with market data

- We lack sufficient price variation in our data (needed to seperately identify premium coefficients from price sensitivity) and the estimated preference distribution strongly supports negative premium coefficients

  $\rightarrow$ marginal cost estimates that seem unrealistically low (margins $> 100\%$)

- Our approach: constrain quality tiers within a brand:

$$\beta_{JackDaniel's}^{base} \leq \beta_{JackDaniel's}^{premium}, \tag{4}$$

  to interpret households purchasing base quality as being price sensitive instead of disliking higher quality whisky

- **Note**: this does not constrain horizontal preferences among brands, e.g. $\beta_{JackDaniel's}^{base}$ compared with $\beta_{JimBeam}^{base}$

MCMC sampler

- The Bayesian implementation of our demand model follows the approach in Pachali, Kurz and Otter (2018)
- Unconstrained coefficients have a standard normal prior while sign and order constraints are imposed through a log-normal distribution
- MCMC inference is performed on a transformed space, such that coefficients are jointly normally distributed after the transformation
- Specifically, we define the function $g : \mathbb{R}^k \rightarrow \mathbb{R}^k_c$ mapping conditionally normally distributed variates θ_i^* to sign and order constrained coefficients θ_i that enter the likelihood

  Note: mapping g needs to be manually adjusted from case to case

Specification of subjective priors

- The implementation in Pachali, Kurz and Otter (2018) allows to specifiy subjective priors of unconstrained and constrained coefficients separately from each other
- This is necessary as the two represent distinct distributions on the transformed θ-space (normal vs. log-normal distribution)
- By default, we use standard weakly informative subjective priors for the parameters entering the hierarchical prior of unconstrained coefficients
- The log-transformation we employ to draw from the posterior of constrained coefficients requires somewhat "tighter" subjective priors

  $\rightarrow$ avoids high prior variance on the transformed space
- Our methodology thus allows for the combination of thight priors on coefficients to be transformed exponentially and a diffuse prior on unconstrained coefficients

Discussion

- We illustrated the benefits of sign and order constraints for structural parameter estimates and implied counterfactual predictions
- Pior constraints rule out model explanations of the data that violate common market knowledge and basic economic theory (e.g., "low-income households dislike premium products")

  $\rightarrow$ interprets households as being price sensitive when purchasing low quality products
- We document that the presence of economically motivated constraints in the hierarchical prior substantially improves face validity of marginal cost estimates
- Sign and order constraints are also crucial in cases where sufficient price variation is present (e.g., Conjoint applications where individual level data is small)

Thank you for your attention!