ST308 Final Project

28545

Introduction

In this study we will be analyzing data from a choice based conjoint study ran on 250 potential customers in February 2021. Conjoint analysis is a state-of-the-art marketing tool designed to discover which features of a given products (and which combinations of such features) customers value the most. One of the important aspects of conjoint analysis is that it incorporates potential prices into customers decisions and as we find out in this study can also incorporate how demographic features of respondents impact those decisions. These aspects make it the perfect study for deciding the design of our product as well as it's future price.

Conjoint Analysis overview

Conjoint studies are very simple in principle. In preparation of the study we first need to think of what features of our product we want to study (we often call those the attributes). For example we can think of the study ran by [1] Nils Goeken, Peter Kurz, and Winfried J. Steiner, on tire purchases. The attributes studied there would be brand, tire type, longevity, rolling resistance, grip on wet roads, customer reviews, independent study results, as well as a price. One of the strong sides of the conjoint analysis is that we can not only test product's features themselves but also importance of marketing specific features such as brand or customer reviews. Next we need to decide on the numerical quality of the features, which we call attribute levels. For example for customer reviews this would be level 1: reviews not available, level 2: 1 star, level 3: 2 start and so on… The same is of course true for the price which should incorporate multiple levels for reliable results.

Once we have our attributes we now need to prepare our product cards. These are simple representations of our products with description of what features a given combination contains as well as the price. Preferably, for maximum reliability, we would also want to include pictures of the given products (or even for the features) as customers often make choices based on visual cues. We of course need to be careful when giving visual representation as for some products (such as a smartphones), look of the product has been proven to influence customers' decisions, hence it is worth thinking whether we would not prefer to include the look of a product as a distinct attribute.

So how do we actually run the study? Every customer is presented with multiple rounds of questions in which they're presented with the product cards and are asked to either rank most to least likely to buy, or choose a single product they'd want to buy. The latter is called choice based conjoint, and is the survey we have ran in this project. Example below:

The amount of rounds should reflect the amount of cards, meaning every customer should see every possible card at least once. To avoid any possible biases of survey creators, at each round the choices presented to the customer should be randomized. Luckily all of this can be done by most modern survey software, and we can focus on the data analysis.

Dataset

We will be using 2 dataset, CBC_data and Demographic_data. The first dataset contains only the output of Conjoint part of the study. Demographic dataset has 3 columns of demographic features. We coded gender

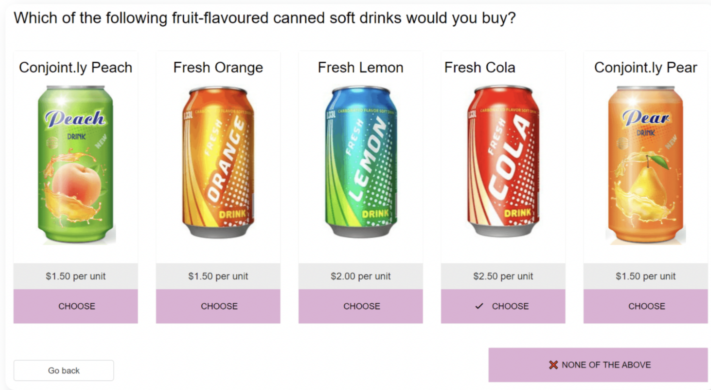


Figure 1: Example of conjoint survey from the customer point of view

as a binary variable, 1 if the responded is female and 0 if the responded was male (no respondents indicated other genders). Children column indicates how many children the respondent had which we encoded by an integer. Competition column indicates whether our respondents know of any similar products (the main subject of our survey is an innovative product, hence we might have reasons to believe this might have some impact).

CBC (choice based conjoint) dataset columns 4-9 are binary variables (excluding price) indicating what set of questions contained which features of the product (encoded 1 if a given feature was containted in a given card). 'ID' column indicates the ID of the respondent, of which we sub sampled 100 (due to computational constraints). Each respondent answered 8 rounds of conjoint, number of the round indicated by 'Set' column, in which each respondent was presented with 4 different choices, indicated by 'Card' column. Last column indicated an independent binary variable for a chosen product (1 - chosen set, 0 - not chosen).

For example 11th row indicates, 1st respondent's, 3rd round of questions', 3rd card, which contained Product 1, Feature 2 and 3, at a price of £7.99, which was chosen out of the remaining 3 other product cards in that round.

It is worth noting that we ran the survey with a 'NONE' choice, giving our customers an option to indicate they would be not interested in purchasing any of the presented combinations of products. This is simply encoded as a row of all 0's (including price), and has a card number 4 in each round. Giving respondents an option to not purchase any products is a good practice when running choice based type of conjoint [2] as it allows us to better estimate how many of targeted demographics are actually our potential customers.

Dataset is solely owned by the author of the study. Survey respondents have been provided by a 3rd party and all identifying features of either the product or respondents have been removed in the process of data cleaning for the purpose of anonymity. The data used in this project is a sample of 100 respondents out of the entire survey for the purpose of limiting computational power required

Bayesian Approach

There are multiple benefits to running Bayesian approach. Firstly, we can incorporate previous beliefs about given parameters. This can come in handy when our company has previous experiences with selling similar products and we can estimate the increase in likeliness of purchase based on previous experiences of managers. Conjoint studies actually have a specific methods of incorporating managers views into our analysis [3]. But even when we are not incorporating any specific information and just using diffuse priors (like we will in this study), there are certain benefits of using Bayesian methods.

More importantly Bayesian models give us an estimate of the entire posterior distribution as opposed to a maximum likelihood point estimate. Now, thanks to central limit theorem we know that, summed up independent draws from any distribution will eventually tend to a normal distribution[4]. This means the posterior of models, such as logistic, despite the fact that they rely on Bernoulli distribution, will eventually converge to a normal. This effect is known as an asymptotic convergence and allows us to estimate the posterior just using the frequentist models. Sadly, the amount of samples for which the asymptotic convergence occurs is often hard to predict thus when an accurate prediction of a posterior is of importance to us, and we cannot be certain we have enough samples, Bayesian models should be applied[5]. This is exactly the setting of conjoint analysis, and even more so of the Hierarchical part of Conjoint studies.

The outputs of conjoint analysis are often used for forecasting revenues or marketing decisions, which give special importance to accurate posterior. For example, assuming that our real posterior of the price looks more like a t-distribution (more weight on the tails), trying to create a pricing strategy using the frequentist approach might make us set a too low initial price and give a too low discount (believing that the distribution around the mean of the price is steeper than it actually is).

Furthermore, in order to be able to target our audience effectively we need to address our customers' differing needs and wants (customer heterogeneity). This is exactly why we would want to use a Hierarchical model[1], treating each survey respondent separately, with group-level coefficients for demographics. Given we only have 32 survey responses per person we can expect our posterior estimates to vary significantly from the normal distribution.

But there is yet another reason why we will run a Hierarchical Bayesian model. In a standard frequentist approach we would either be pooling the entire sample assuming every group is the same or analyze them independently (every group is independent of any other). Bayesian Hierarchical model allows us to choose the degree of pooling our observations. Using HB (Hierarchical Bayes) we can analyze differences in groups assuming that as a whole they do still follow some general distribution[9]. In our study we can can be certain despite the differences in tastes of individuals some patterns will follow (for example we expect the predictor for the price to be negative for all respondents). Being able to address both the differences in tastes and the purchasing patterns of the entire population should give us a much results. To quote Nils Goeken, et al [1] "There is strong empirical evidence that addressing individual preference heterogeneity in CBC studies using HB (Hierarchical Bayes) modelling pays off in terms of statistical model performance (e.g. providing a higher forecasting accuracy)...".

Model Explanation

Under Choice Based conjoint our customers are only making one choice (like in real life setting) of either buying a given product or not. We can see this as a classification problem with a binary independent variable. Firstly, we will ignore the conditional nature of people's choices and run a simple Bayesian logistic regression, where our coefficients will be the corresponding features/ products.

$$log(\frac{\pi(x)}{1-\pi(x)}) = \beta_1 x_{product_1} + \beta_2 x_{p_2} + \beta_3 x_{feature_1} + \beta_4 x_{f_2} + \beta_5 x_{f_3} + \beta_6 x_{price}$$

It is worth noting that we are not including a constant term. This is because the 'NONE' option is indicated by a row of all 0's and we cannot rationally expect customers to donate money to us willingly.

Next we will run a Hierarchical model. In order to be able to incorporate group-level predictors we will also need to extend our Bayesian logistic regression to a multinomial conditional logistic regression. Our model will look in the following way [6] (where X indicates a matrix and β_j a vector of attribute coefficients for the j'th respondent):

$$\eta_{ij} = z_{gender}\gamma_1 + z_{kids}\gamma_2 + z_{competition}\gamma_3 + X_{ij} * \beta_j$$

For the priors we are going to use diffuse priors as we don't have any previous data or anyone's opinions to elicit. We will be using the standard "Mixture-of-Normals" priors often used in similar studies [1][7] and explained in-depth in STAN user's guide [8].

Bayesian Logit Model

We will run 4 chains with 100,000 iterations (with a standard half of the sample burn-in).

Before examining the results let's visually check the convergence:

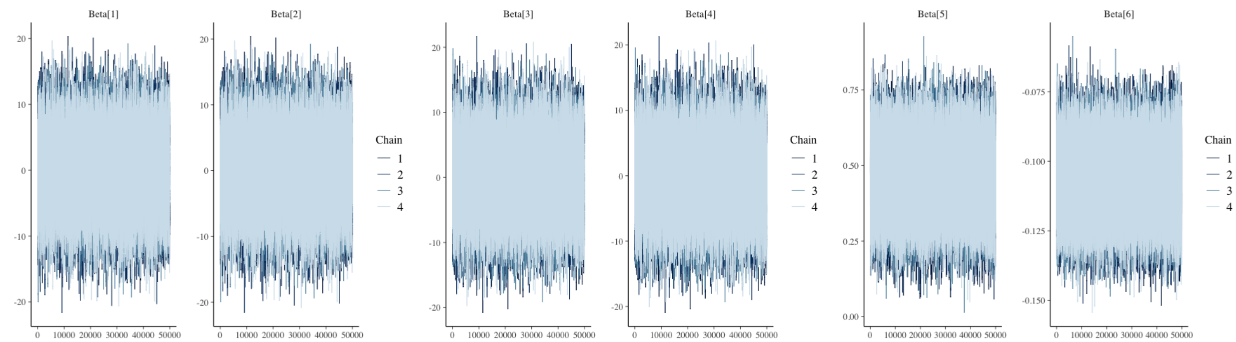


Figure 2: Traceplots for Bayesian Logit

We can see from the plots that all of our chains are steadily sampling from a limited range of values. This tells us they have converged around what we believe to be the true posterior, irrespective from what values the chains have started. We can thus expect our posterior distributions to give us good estimates.

Let us now plot the posterior distributions with 75% credible interval:

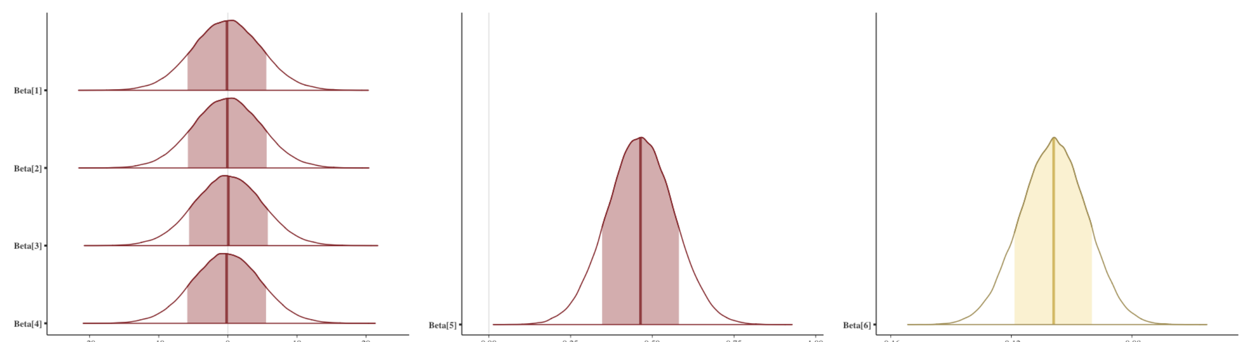


Figure 3: Posterior distributions

First thing we can immediately notice is our β 1-4 coefficients have long tails, as compared to Feature_3 (β_5) and a price predictor (graphed in yellow) which are narrowly concentrated around their means. This could suggest our respondents have widely varying preferences for attributes 1-4, something we will address with a Hierarchical model in a moment.

Another point worth noting is that the posterior of the price (yellow) lies entirely below 0. This is exactly what we would expect. The higher the price, the lower the probability a customer purchases our product.

But in order to appropriately interpret the coefficients we need to look at the precise values. Let's look at the means and credible interval in more detail (Figure.4).

We can see that the mean of the output is in fact close to but not actually 0. This is most likely cause by a narrow diffuse prior with $\sigma = 10$. Nonetheless, this should not be a problem for us.

So how do we actually interpret the coefficients? Firstly, we can see that the means for Product_1 / 2 and Feature_2 are negative. This might seem like an odd result but it just tells us that an average person would not be willing to purchase either the product 1 or product 2 alone. What about a mix of products

```
Inference for Stan model: CBC_Bayesian_logistic_regression.
4 chains, each with iter=1e+05; warmup=50000; thin=1;
post-warmup draws per chain=50000, total post-warmup draws=2e+05.

            mean se_mean   sd     2.5%      25%      50%      75%    97.5% n_eff Rhat
Beta[1]    -0.13    0.02 4.97    -9.89    -3.48    -0.13     3.22     9.64 43241    1
Beta[2]    -0.08    0.02 4.97    -9.84    -3.43    -0.07     3.27     9.69 43245    1
Beta[3]     0.08    0.02 4.97    -9.69    -3.27     0.07     3.43     9.83 43248    1
Beta[4]    -0.18    0.02 4.97    -9.95    -3.53    -0.19     3.17     9.59 43233    1
Beta[5]     0.46    0.00 0.10     0.26     0.40     0.46     0.53     0.66 85451    1
Beta[6]    -0.11    0.00 0.01    -0.13    -0.11    -0.11    -0.10    -0.08 88907    1
lp__    -1754.47    0.01 1.72 -1758.66 -1755.39 -1754.15 -1753.21 -1752.10 63656    1
```

Figure 4: Example of conjoint survey from the customer point of view

and features. We can do quick calculus for how likely an average customer would be willing to purchase a Product_2 (β_2) with Feature_3 (β_5) at a £5 price:

$$\frac{e^{-0.07+0.46-0.11*5}}{1+e^{-0.07+0.46-0.11*5}} = 0.46$$

There is around 46% chance one of our respondents would make a purchase for such a product.

We can actually reverse the calculation and calculate what would be the highest price our customer would be indifferent between purchasing and not purchasing (50%) for a product combination. The log-odds of 50% probability is exactly 0 so we're looking for such a price for which vector $X * \beta = 0$:

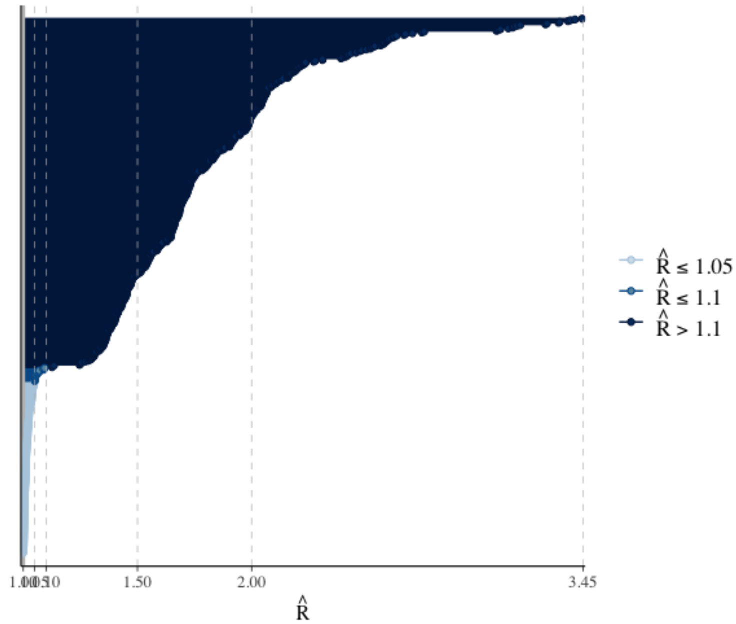
$$\frac{-0.07+0.46}{0.11} = 3.54$$

We can see our average respondent would be willing to pay £3.54. At a 75% credible interval values this price would go up to £37.5, implying 25% of our customers might be willing to pays such a price. We could use the entire posterior to draw a demand curve using this technique, but here is where we would be making an important (and unlikely to be true) assumption, that our customers' tastes will be the same for the entire population (we are pooling all of the respondents' answers).

To solve this problem and be able to predict true demand curve we should apply the Hierarchical model.

Hierarchical Model

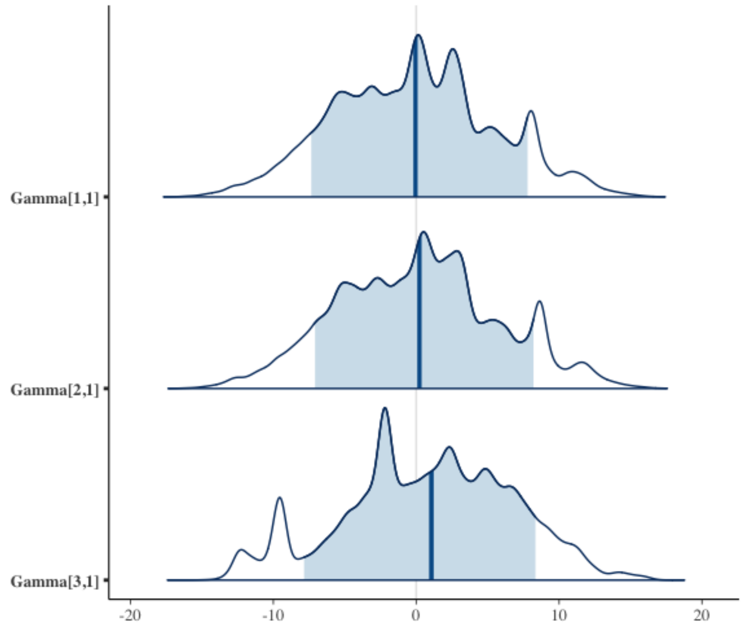
Our Hierarchical model has over 600 parameters, hence we can no longer check the convergence manually. Instead let us try to plot the rhat values.



Rhat value is one of the most popular statistic for testing convergence. Essentially it compares the means and variances of our chains to decide whether the chains have converged to a stable point and started sampling from where we believe our posterior to center around. A good sign of convergence is rhat value of below 1.05 and above 0.99.

We can clearly see from the picture that a large portion of our coefficients has values significantly above 1.05, implying our chains give widely varying samples. This is rather surprising given we have ran it on 100,000 iterations which should be more than sufficient for such studies. One solution would be possibly running more iterations but this would require computational power beyond that allocated for this study. For now we will proceed with our coefficients as they are although,from the poor convergence we can expect our posterior distributions to be skewed.

Let us first examine a couple of coefficients for the group-level predictors we have added. These are the posteriors for the effect of Gender on differences in preferences for product 1, product 2 and feature 1:



We can see that Gender on average doesn't really seem to have a lot of impact on the preferences between Product 1 and 2 but does seem to have a positive impact on preference for Feature 1. Again, the scaling of the coefficients might give us the idea that the results are insignificant but the coefficients should be considered in relation to the price coefficient (the scaling shouldn't matter too much in this case).

It is also worth noting the weird shape of the posterior with multiple kinks. These weird kinks could be the effect of our mcmc chains sampling unevenly from the posterior.

Now that we have ran a Hierarchical model, we can actually compare individual preferences of respondents. As an example let's take respondent 1 and 2:

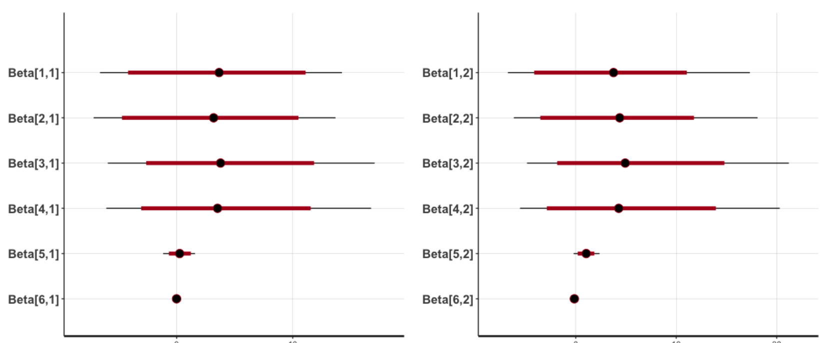


Figure 5: Coefficients of Respondent 1 (left) and Respondent 2 (right)

Comparing the means of responses we can see individual 1 prefers Product 1 over product 2 whereas respondent 2 prefers the reverse. Respondent 2 also seems to value Feature 1 (β_3), much more than respondent 1 (relative to Products 1 & 2). Despite the slight differences in means we should be careful about making any judgments given the very wide credible intervals (which have also came up in the pooled model). We can see that, Betas 1-4 have very wide credible intervals as compared to Beta 5 and 6. One explanation might be that our customers actually struggle to choose themselves. The conjoint study concerns a rather innovative previously unseen product, which might make customers uncertain of what to pick. The poor convergence of our model can also in part explain the large confidence intervals.

But how different are the tastes of our respondents? We can analyze the price coefficient, for the first 50 respondents to see the between group variation.

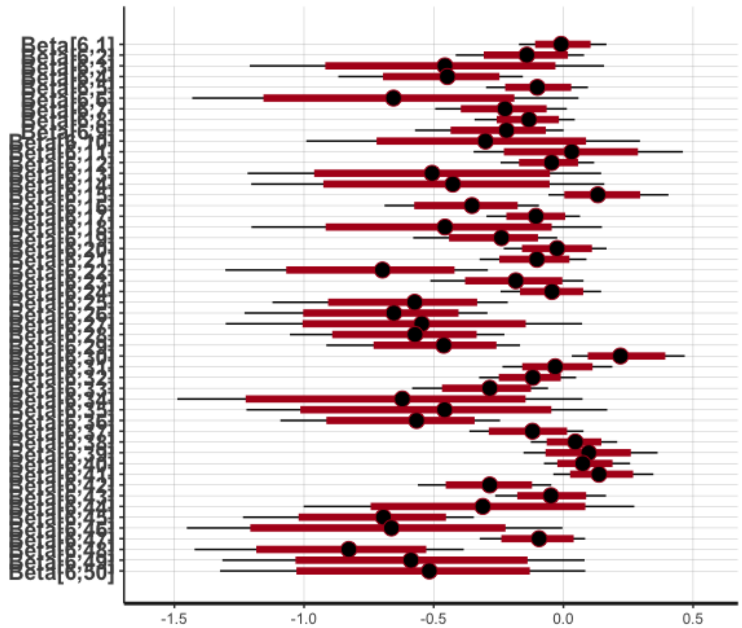


Figure 6: Effect of the price coeffcient for the first 50 respondents

As expected our Beta for price is small, negative and in many cases significantly different from 0. Despite the fact that the pool model was very concentrated around -0.11 our respondent's sensitivity to price differs very significantly from an individual to individual.

One interesting observation is that some of our estimates actually lie above the 0 value. This would imply an improbable scenario that those customers actually like paying a higher price for a product?! Although, lack of convergence might not be helping with the credible intervals it certainly isn't causing this weird result. Here, we're most likely encountering a human error, or more specifically survey fatigue. Our survey took around 9 minutes to complete (which is a lot) so we can expect some respondents to start clicking random answers. One advantage of Hierarchical model is we can better identify such abnormal responses, and exclude them when forecasting for much better results.

References

[1] https://bms-net.de/wp-content/uploads/2021/11/Hierarchical-Bayes-Conjoint-Choice-Models-\T1\textendash-Model-Framework-Bayesian-Inference-Model-Selection-and-Interpretation-of-Estimation-Results-Goeken-N.-Kurz-P.-Steiner-W.-2021.pdf [2] https://eml.berkeley.edu/~train/foundations.pdf [3]https://www.jstor.org/stable/3152044?seq=4 [4]https://en.wikipedia.org/wiki/Central_limit_theorem [5]https://stats.stackexchange.com/questions/296220/why-i-should-use-bayesian-inference-with-uninformative-prior [6]https://data.princeton.edu/wws509/notes/c6s3 [7]https://rpubs.com/TCEagle/672172 [8]https://mc-stan.org/docs/2_18/stan-users-guide/multivariate-hierarchical-priors-section.html [9]https://www.bayesrulesbook.com/chapter-15.html#partial-pooling-with-hierarchical-models