Bayes Sampler Assignment

Jan-Willem Simons
E-mail: j.g.simons@uu.nl

Research question

An important topic in economics and sociology relates to the issue of how markets decide on product prices. One hypothesis in economics posits that the objective quality of a product is the sole determinant of its market price. For example, suppliers and buyers might use quality indicators like clock speed in megahertz (MHz), hard drive size in megabytes (MB), and random-access memory size (RAM) in MB to determine the price of a personal computer (PC). Sociological theory would conversely posit that markets do not decide prices on the basis of product quality alone but will also valuate the buyer's normative perceptions about the product. Examples of normative perceptions which influence price are the perceived status that the product provides to the buyer, or how advertisements increase the products' perceived attractiveness. Thus, under the sociological framework, not only the quality of the PC, but also whether it was manufactured by Apple, or whether it was heavily advertised, would determine its price. Although the importance of objective price determinants can hardly be questioned, that of normative price determinants can. Research by Anselmsson, Bondesson and Johansson (2014) found that brand equity led to a price premium on grocery products in Sweden, whereas Holbrook (1992) found that brand equity did not result in price premiums in the electronics sector in the US. This study seeks to add to the subject by investigating the degree to which computer prices in the US in the early 90's were determined by normative price determinants above and beyond that of objective determinants. The following research question is formulated: controlling for objective price determinants, did normative price determinants influence the price of personal computers in the United States during the 1993 – 1995 period?

Introduction of the dataset

The dataset that will be used to investigate the research question is the "computers" dataset in the R-package "Ecdat", which was originally collected by Stengos Zacharias (2006). It consists of a total of 6259 observations on the advertised price and features of PC's, which were collected from a leading PC magazine in the US on the first of every month from

January 1995 to November 1995. Besides the advertised price, five features of the observed PC's were recorded. These were the clock speed of the PC in Mhz, the size of its hard drive in MB, its RAM size in MB, the total number of advertisements, and a dichotomous variable indicating whether the manufacturer of the PC was a premium firm or not. The first three variables are used to represent the objective price determinant, whereas the final two are used to represent the normative price determinant. See table 1 for descriptive statistics of the dependent variable and the grand mean centered independent variables. Two multiple linear regression models are consequently specified to represent the competing theoretical models of interest, one where the advertised price is regressed on the objective price determinants only, and one where the advertised price is regressed on both the objective and normative price determinants. The efficacy of these two models will be compared by means of the DIC. Bayes factors will be used to evaluate hypotheses for the best fitting model.

Table 1

Descriptives of the outcome variable price and its grand mean centred determinants speed, HD, RAM, ads, and premium.

Variable	N	Mean	Std. Dev.	Min.	Max.
Price	6259	2219.58	580.80	949.00	5399.00
Speed	6259	0	21.16	-27.01	47.99
HD	6259	0	258.55	-336.60	1683.40
RAM	6259	0	5.63	-6.29	23.71
Ads	6259	0	74.84	-182.30	117.70
Premium	6259	0	0.30	-0.90	0.098

Estimation – Gibbs sampling & the Metropolis-Hastings algorithm

A Gibbs sampler is used to sample from the conditional posteriors of the determinants "speed", "HD", and "RAM" in both models. The conditional posteriors of these three parameters are derived by combining a normal data likelihood with a normal informative prior. A normal prior is used because it is a conjugate prior for the likelihood, giving a closed-form expression for the posterior. Each normal prior is assigned uninformative values for its mean and variance because unfortunately no suitable values were provided in Stengos Zacharias (2006). In both the first and second model, the conditional posterior of the residual variance is obtained by combining the normal data likelihood with an uninformative conjugate inverse-gamma prior, resulting in an inverse-gamma posterior that is bounded on the interval from zero to infinity. No information about the residual variance is provided in the article by Stengos Zacharias (2006) either, so that no informative values for the inverse-gamma residual variance parameter can be specified.

A Metropolis-Hastings (MH) step is consequently used to sample from the conditional posteriors of the normative price determinants "ads" and "premium" in the second model. Since the sample was obtained from a single PC magazine in the US, it might not be representative of the entire population of PC listings in the US. Stengos Zacharias (2006) similarly note that their sample might not be fully representative of PC listings in the US, because they only collected information on the prices and features of PC's from a limited subset of manufacturers, and for example did not sample PC's listed by computer wholesalers. In order to account for potential sampling induced bias in the conditional posteriors of the determinants "ads" and "premium", these posteriors are derived by combining a normal data likelihood with a t-prior. The t-prior is specified such that it introduces a high degree of uncertainty in the conditional posteriors of the "ads" and "premium" variables, namely by assigning it a mean of zero, a variance of 0.001, and one degree of freedom. Since the resulting conditional posteriors are intractable, an MH step is required to sample from them. Normal distributions with mean and variance ML estimates of the determinants "ads" and "premium" are used as independent, fixed proposal distributions for doing so. This study opts for an approximation of the marginal posterior which does not have to be evaluated in each iteration, because the procedure would otherwise be too time consuming due to the associated computational intensity of finding a well approximating density. The variance of the proposal distributions is furthermore multiplied by a factor of 2.5, because trial and error indicated that the proposal distributions would otherwise be concentrated too tightly around the MLE estimate of the coefficients of the determinants, so that no uncertainty would be introduced in their respective conditional distributions. Finally, although it can be argued that an MH step needs to be used to account for sampling bias in each of the determinants in the model, here it is only used to sample from the "ads" and "premium" determinants. The reason is that these two variables are of primary substantive interest, so that estimates from their conditional posteriors should be as precise as possible.

An overview of the means and variances of the various priors in each of the two models is given in table 2 on the previous page. In both models, two chains with a length of 12.000 iterations, each with a burn-in of 2.000, for a total of 10.000 iterations per chain are used for approximating the conditional posteriors of the relevant parameters. Generic initial parameter values are used to initialize the chains in each model. Point estimates from the conditional posteriors are obtained with expected a posteriori (EAP) estimates and 95% credible intervals (CI's). All independent variables were centered to reduce auto-correlation and facilitate convergence throughout the parameter space.

Table 2

Prior specifications for the parameters in both models.

Model	Parameter	Method	Prior	Prior mean	Prior variance
Model 1	Intercept	Gibbs	Normal	0	1000
	Speed	Gibbs	Normal	0	1000
	HD	Gibbs	Normal	0	1000
	RAM	Gibbs	Normal	0	1000
	Residual variance	Gibbs	Inverse gamma	0.001	0.001
Model 2	Intercept	Gibbs	Normal	0	1000
	Speed	Gibbs	Normal	0	1000
	HD	Gibbs	Normal	0	1000
	RAM	Gibbs	Normal	0	1000
	Ads	MH	t	0	0.001
	Premium	MH	t	0	0.001
	Residual variance	Gibbs	Inverse gamma	0.001	0.001

Assessing model convergence

For both models, model convergence is assessed by inspecting history plots, autocorrelation plots, and Markov chain errors, while for the second model the acceptance ratio of the Metropolis-Hastings algorithm is additionally assessed. The history plots show no indication of non-convergence for any of the parameters in either of the two models, meaning that for each parameter the two chains are visually superimposed on top of each other. As such, it seems to be the case that the Gibbs sampler and the Metropolis-Hastings algorithm have adequately converged through the conditional posterior parameter spaces of both models, with the caveat that one can never be certain whether chains have converged since they are asymptotic approximations.

The autocorrelation plots furthermore show that in both models there is some autocorrelation between iterations in the chains of the "RAM" and "HD" parameters, but almost none in the chains of the intercept, the "speed" parameter, and the residual variance. More specifically, the "RAM" and "HD" parameters have an autocorrelation of lag five, where in the second model, the parameters "ads" and "premium" additionally show an autocorrelation of lag four. These results imply that both the Gibbs sampler and the MH algorithm have some difficulty converging through the parameter spaces of the "RAM", "HD", "ads", and "premium" coefficients. Besides centering, one solution to this issue could be to derive the conditional posterior distributions for these parameters on the basis of a thinned chain, meaning that iterations in the chains would be uncorrelated if the chains would be thinned such that they contained only each fourth or fifth iteration. This study chooses not to adopt this method because the utility in terms of precision would be marginal while the added computational burden would be significant. Since the autocorrelation values are not substantial to the degree that they are indicative of serious non-convergence, they are considered to be acceptable.

The Markov chain error of each parameter in each model is obtained by dividing the naïve standard error over the combined iterations of the two chains by the square root of 20.000, which is the total number of iterations over the two chains. It is consequently assessed whether any of the Markov Chain Monte Carlo (MCMC) errors are larger than 5% of the naïve standard error, which would be an indication of substantial error in the chain due to sampling. The MCMC error is found to be smaller than 5% of the naïve standard error for each parameter in both models, meaning that the error due to MCMC sampling is negligible. The specified number of iterations per chain thus seems to be sufficient, with the MC error values providing no evidence of model non-convergence.

The acceptance ratio of the MH algorithm is finally 42.18% for the "ads" parameter, and 44.43% for the "premium" parameter. Although these percentages are relatively low for an independent proposal distributional, and would thus be indicative of slow convergence, they are partly a product of the fact that the variance of the proposal distribution is multiplied by a factor of 2.5 to introduce additional uncertainty into the estimates for the "ads" and "premium" parameters. Without that correction, the acceptance ratio for both parameters would be around 75%. As such, in order to introduce uncertainty into the conditional posteriors of "ads" and "premium", concessions in the efficiency of the MH algorithm unfortunately have to be made. For the purposes of this study, the convergence rate of the MH algorithm is considered to be acceptable.

Posterior predictive checks

Posterior predictive checks (PPC's) are used to evaluate two linear regression assumptions: homogeneity of variance and normality. The following procedure is used to determine the posterior predictive p-value for each assumption: first, discrepancy measures are chosen that can be used to evaluate whether a set of residuals show homogeneity of variance or are normally distributed. The ratio of the variance of two separate but equal halves of the observed data is used as a discrepancy measure for evaluating the homogeneity of variance assumption, while the skewness of the distribution of the residuals is used as a discrepancy measure for evaluating the normality assumption. The first discrepancy measure has a ratio of one when the variances in the two separate data halves are equal, has a ratio larger than one when the variance of the halve in the numerator is larger, and has a ratio below one but above zero when the variance of the halve in the denominator is larger. The second discrepancy measure has a value of zero when the residual distribution is perfectly symmetric, a negative value when the distribution is right-skewed, and a positive value when the distribution is left-skewed. For both discrepancy measures, extreme values in either direction are indicative of strong departures from the assumption of interest. The second step in the procedure is to use the set of model parameter estimates in each MCMC iteration to calculate residuals for the observed data. These parameter sets can consequently also be used to generate replicated datasets, for which residuals can similarly be calculated. Here, the number of unique parameter sets is equal to the number of iterations in the combined chains, i.e., 20.000 per model. The third step is to calculate the discrepancy measure over each of the observed and replicated residual sets. The resulting 20.000 discrepancy measures per observed and replicated residual set under each model can then be used to calculate the Bayesian p-value. This can be done by assessing how many of the iteration-ordered replicated discrepancy measures are larger than the iteration-ordered observed discrepancy measures. Very high or low PPC values indicate that the observed data do not resemble the posterior predictive distribution, meaning that the null-hypothesis that the model conforms to the assumption of interest should be rejected. In the first model, the PPC for the homogeneity of variance assumption is equal to 0.5057 where the PPC for the normality assumption is equal to 0. In the second model, the PPC for the homogeneity of variance assumption is equal to 0.5013 where the PPC for the normality assumption is equal to 0. It can thus be concluded that homogeneity of variance holds in both models, whereas normality does not.

Model selection using the DIC

The deviance information criterion (DIC) is used to assess and compare the relative fit of the first model with the objective PC features to that of the second model which also includes the normative PC features. Comparing DIC values for these two competing models will provide a measure of whether the increased model complexity that results from adding the two normative price determinants is offset by the increase that their inclusion offers in terms of model fit. This comparison consequently enables the identification of the superior model in terms of model fit, and will provide an initial indication of the degree to which computer prices in the US in the early 90's were determined by normative price determinants.

In short, the DIC expresses model fit as a function of model fit and model complexity. It does so on the basis of the following formula: $DIC = -2logf(y|\hat{\theta_y}) + 2p_D$. The first term in the formula is the model misfit, and is calculated as the likelihood evaluated at the posterior mean of θ. The second term is an estimate of the effective number of parameters, i.e., the complexity of the model, and is calculated by subtracting the likelihood evaluated at the posterior mean of θ from the average likelihood over the posterior distribution of θ, and multiplying the result by two. The DIC is then equal to the sum of these two respective terms. The mean deviance is consequently equal to 94023 in the first model and 93205 in the second model. The penalty term is equal to 62 in the first model, where it is equal to 59 in the second model. As such, the DIC or penalized mean deviance is equal to 94085 in the first model and 93263 in the second model. A difference in DIC of $94085 - 93264 = 821$ is observed between the first and second model.

Based on the rule of thumb that a difference of 10 indicates that the model with the higher DIC can be ruled out, it can convincingly be concluded that the model fit of the second model is superior to that of the first model. The second model is thus selected for interpretation and hypothesis evaluation with the Bayes factor, and the tentative conclusion is drawn that computer prices in the US in the early 90's were in part determined by normative price determinants.

Model selection using the Bayes factor

Three hypotheses are a-priori formulated and assessed by means of Bayes factors. The first two hypotheses relate to the sign and strength of the effect of the price determinants. The effect of each price determinant can reasonably be expected to be positive. Manufacturers are likely to demand a higher price for a higher quality PC, whereas increases in the number of advertisements and a manufacturers' status are likely to increase the demand for a PC and its price as well. The first hypothesis therefore states that: H1: The effect of each of the objective and normative price determinants is positive. Second, since Stengos Zacharias (2006) chose their determinants on the basis of theoretical expectations that these would affect price, it seems reasonable to assume that each determinant will show a minimally small effect in terms of Cohen's D. The second hypothesis thus states that: H2: Each of the objective and normative price determinants has a standardized effect of at least 0.1. The third and final hypothesis relates to the difference in the effect of the respective price determinant types. Since the literature generally indicates objective price determinants to be more important than normative price determinants, the effect of each of the objective price determinants is expected to be stronger than that of each of the normative price determinants. The third hypothesis consequently states that: H3: Each of the objective price determinants has a larger effect on PC price than each of the normative price determinants.

The Bayes factor for each hypothesis is consequently determined by dividing the fit of the model under the hypothesis by the complexity of the model under the hypothesis. The fit for each hypothesis is obtained by approximating the joint posterior distribution of the model parameters, and evaluating it under the hypothesis of interest. The joint posterior distribution is approximated by drawing 10.000 values from a multivariate normal distribution, with a vector of standardized ML parameter estimates as its mean, and the covariance matrix of the standardized ML parameter estimates as its covariance matrix, which is then multiplied by a standard uninformative prior. The complexity for each hypothesis is obtained by approximating the prior distribution of the model parameters, and evaluating it under the hypotheses of interest. The prior distribution is approximated by drawing 10.000 values from a multivariate normal distribution, with a vector of zeroes as its mean, and the covariance matrix of the standardized ML parameter estimates divided by a fraction b as its covariance matrix, where b is equal to 5 / 6259. A vector of zeroes is provided for the mean because the prior distribution needs to be adjusted such that its means are located on the boundaries of the hypothesis under consideration. The covariance matrix is divided by the fraction b to obtain the prior distribution of the parameters, where the value for b is chosen arbitrarily, and is here based on the number of determinants divided by the sample size. The fit and complexity are consequently divided to obtain Bayes factors of 0 for the first hypothesis, 0 for the second hypothesis, and 0.017 for the third hypothesis. As such, there is no support for the first and second hypotheses, while there is almost no support for the third hypothesis. It can be concluded that not all of the effects of the price determinant are positive, that some effects have a standardized effect smaller than 0.1, and that each of the objective price determinants is not more important than each of the normative price determinants.

Interpretation of estimates and intervals

As stated earlier, the results of the analysis are assessed by calculating expected a posteriori (EAP) estimates and 95% credible intervals (CI's) of the conditional posterior distributions of the parameters of interest, here those in the second model. The EAP estimates and their associated CI's are provided in table 3 on the next page.

As hypothesized, the determinant "speed" is positively related to "price", where for every Mhz increase in clock speed, the price of a PC is predicted to increase by 6.56 dollars, with a 95% certainty that the value for the coefficient lies in the [6.36; 6.75] interval. The "HD" determinant is negatively related to "price", where for every hard drive MB increase, the price of a PC is predicted to decrease by 0.34 dollars, with a 95% certainty that the value for the coefficient lies on the [-0.38; -0.31] interval. This is a counter-intuitive finding, and not in line with what was hypothesized. One possible explanation for this finding could be that high-end PC's in the 90's had to sacrifice hard drive space for higher performance. As hypothesized, the "RAM" determinant is positively related to "price", where for every RAM point increase, the price of a PC is predicted to increase by 77.80 dollars, with a 95% certainty that the value of the coefficient lies in the [75.67; 79.97] interval. As hypothesized, the "ads" determinant is positively related to "price", where for every increase in the number of ads, the price of a PC is predicted to increase by 1.26 dollars, with a 95% certainty that the value of the coefficient lies in the [1.13; 1.39] interval. Finally, if the PC manufacturer is a premium firm, the price of a PC is predicted to decrease by -394.44 dollars, with a 95% certainty that the value of the coefficient lies on the [-427.44; -361.53] interval. This is not in line with what was hypothesized. A possible explanation is that premium firms can afford to lower prices due to high turnover. Based on these results and the value of the Bayes factor for the third hypothesis, it can finally be concluded that normative price determinants clearly influenced the price of personal computers in the United States during the 1993 – 1995 period, above and beyond the effect of objective price determinants.

Table 3

EAP estimates and 95% CI's of the posteriors of the parameters in the second model.

Parameter	EAP [2.5% C.I.; 97.5% C.I.]
Intercept	2159.78 [2106.24; 2212.17]
Speed	6.56 [6.36; 6.75]
HD	-0.34 [-0.38; -0.31]
RAM	77.80 [75.67; 79.97]
Ads	1.26 [1.13; 1.39]
Premium	-394.44 [-427.44; -361.53]
Residual variance	172556.14 [164727.80; 182731.60]

Comparison of Bayesian and frequentist approaches

One drawback of using the frequentist approach for analysing the "computers" data is that the regression effects would quickly have become significant due to the large sample size. As such, their being significant would not necessarily have been informative. This is not a concern in the Bayesian approach, where regardless of size, the data are incorporated into the posterior with a prior, after which the characteristics of the posterior can simply be evaluated by means of point estimates. Another advantage of the Bayesian approach is that it provides the researcher flexibility for dealing with issues such as sampling bias through use of an MH algorithm, whereas the frequentist method provides no such flexibility. A subsequent drawback of this increased flexibility is that there is more room for researcher induced bias. For example, different posteriors would have resulted if the variance of the fixed proposal distributions would have been multiplied by a factor of

2 instead of 2.5. A third advantage of the Bayesian approach is that Bayes factors can be used to evaluate informative hypotheses, i.e., could be used to directly compare the relative efficacy of objective and normative price determinants. This comparison would not have been as straightforward in a frequent analysis, where classical null-hypothesis testing would need to have been used to arrive at convoluted and mostly unsatisfactory answers. A drawback of the Bayesian approach is that its estimation methods are quite complex and computationally burdensome, especially in comparison to standard MLE regression analysis, which are relatively straightforward to apply and very efficient. With the Bayesian approach the obtained results can finally be incorporated into future research, be it through Bayesian updating or historical priors, which is not possible in the frequentist framework.

References

Anselmsson, J., Bondesson, N. V., Johansson, U. (2014). Brand image and customers' willingness to pay a price premium for food brands. Journal of Product Brand Management.

Holbrook, M. B. (1992). Product quality, attributes, and brand name as determinants of price: The case of consumer electronics. Marketing Letters, 3(1), 71-83.

Stengos, T., Zacharias, E. (2006). Intertemporal pricing and price discrimination: a semiparametric hedonic analysis of the personal computer market. Journal of Applied Econometrics, 21(3), 371-386.