

## Effects of prior distributions: An application to piped water demand

Andrés Ramírez Hassan<sup>a</sup> and Luis Pericchi<sup>b</sup>

<sup>a</sup>*Universidad EAFIT*

<sup>b</sup>*University of Puerto Rico*

**Abstract.** In this paper we analyze the effect on posterior parameter distributions of four possible alternative prior distributions, namely Normal-Inverse Gamma, Normal-Scaled Beta two, Student's t-Inverse Gamma and Student's t-Scaled Beta two. We show the effects of these prior distributions when there is apparently conflict between the sample information and the elicited hyperparameters. In particular, we show that there is not systematic differences of posterior parameter distributions associated with these four priors using data of piped water demand in a linear model with autoregressive errors. To test the hypothesis that this result is due to using a moderate sample size and a relatively high level of expert's uncertainty, we perform some simulation exercises assuming smaller sample sizes and lower expert's uncertainty. We obtain the general same pattern, although Student's t models are slightly less affected by prior information when there is a high level of expert's certainty, and Scaled Beta two models exhibit a higher level of posterior dispersion of the variance parameter.

## 1 Introduction

Despite the fact that there is a debate regarding the relevance of prior robustness analysis; on the one hand the coherent behavior claims for a single prior distribution, but on the other hand it can be very difficult to obtain such a fine prior distribution (Berger, 1985), we think that empirical arguments suggest a combination of elicitation procedures and robustness to possible prior misspecification as an advisable rule. Therefore, the main goal in this paper is to perform posterior sensitivity analysis trying four possible prior alternatives: Normal-Inverse Gamma, Normal-Scaled Beta two, Student's t-Inverse Gamma and Student's t-Scaled Beta two, in an environment where there is apparently misalignment between sample information and elicited expert's knowledge.

We analyze a linear model with autoregressive errors applied to the piped water consumption in the Metropolitan Area of Medellín (Colombia). Additionally, we perform some simulation exercises using smaller sample sizes and different prior covariance matrices that reflect different degrees of expert's certainty to show the dependence of posterior estimates to these characteristics.

The concept of probability from a Bayesian point of view is associated with degrees of belief. In this scenario, the experts' knowledge about an event can be tackled from either a subjective or objective perspective. The construction of prior distributions based on the subjective approach should be adopted in scenarios where it is tenable (Berger, 2006). However, this methodology is strongly influenced by the experts' perception of reality (Garthwaite et al., 2005); unfortunately, experimental exercises have shown that human beings use heuristic strategies to make statistical statements which lead to biased affirmations (Kahneman, 2011). It does not matter which technique is used, the main objective in science is to maximize the process of learning from observation. This observation can be compiled from data and/or researcher's experience. However, what happens when there is a conflict between sample information and prior distributions? Conjugate priors have enormous effects on posterior estimates when there is conflict between data and prior information (Berger, 1994). A possible solution is to use robust priors (Fúquene et al., 2009, 2014). These can handle outliers in a more intelligent way (Bian, 1997), as well as influence in a wiser form the inferential process when there is conflict between prior and sample information (Fúquene et al., 2009). The price to be paid is computational, but nowadays that is not a problem.

In particular, we perform an elicitation procedure with an expert that used to work in the main piped water company of the Metropolitan Area of Medellín (Colombia), and obtain the mean prior elasticities, as well as their variance estimates, associated with the average household consumption of piped water. After we implement an elicitation procedure, we use observed and simulated data to perform sensitivity analysis to prior specifications. We show that posterior parameter estimates are robust to prior distributions, although Student's t priors are less affected by expert's knowledge when there is a high level of certainty regarding prior statements, and Scaled Beta two models show a higher level of dispersion associated with posterior variance.

After this introduction, we outline the principal statements about our model in Section 2.

Section 3 shows the elicitation procedure and its results, and Section 4 exhibits the four models' mathematical specifications. Section 5 shows the principal outcomes of our application, and Section 6 presents some simulation exercises. Finally, we make some concluding remarks in Section 7.

## 2 Bayes regression with autoregressive errors

We study the average household piped water consumption of strata four in the Metropolitan Area of Medellín (Colombia) using quarterly data from 1985 to 2009. Population in this country is divided in strata, the goal is to implement a cross subsidy structure, where strata four pays the reference cost.

We propose a linear model with autoregressive errors due to having time series data with an inertial effect on consumption. Ordinary and partial correlograms indicate an autoregressive process of order one (these outcomes are available upon request). We estimate the following model (Eqs. 2.1 and 2.2)

$$\log \{cme_t\} = \beta_1 \log \{I_t\} + \beta_2 \{n_t\} + \beta_3 \log \{p_t\} + \mu_t \quad (2.1)$$

where

$$\mu_t = \phi \mu_{t-1} + \epsilon_t \quad (2.2)$$

$t = 1, 2, \dots, T$  and  $\epsilon \stackrel{i.i.d}{\sim} (0, \sigma_\epsilon^2)$ .

$\log \{cme_t\}$ : natural logarithm of the average consumption of piped water.

$\log \{I_t\}$ : natural logarithm of average real per capita income.

$n_t$ : average number of people in household.

$\log \{p_t\}$ : natural logarithm of the real price of piped water.

$\mu_t$ : autocorrelated stochastic perturbation.

We must estimate  $\beta_1$  and  $\beta_3$ , which are the income and price demand elasticities, and  $\beta_2$ , which is the semi elasticity of piped water consumption with respect to the number of people in the household. In addition,  $\phi$  captures the inertial effect on consumption, and  $\sigma_\epsilon^2$  is the variance of the random noise.

In particular, we analyze the effects of four prior distributions on the posterior estimates, namely Normal-Inverse Gamma, Normal-Scaled Beta two, Student's t-Inverse Gamma and Student's t-Scaled Beta two (see Section 4 for mathematical details). We assume independent prior hyperparameters because Beach and Swenson (1966) have shown that experts have difficulty giving information about covariance between parameters. In addition, we use as prior for the autoregressive coefficient a truncated normal distribution restricted to the stationary region using as hyperparameters the maximum likelihood estimates, that is, 0.61 and 0.054 are the prior mean and standard deviation, respectively.

The Normal-Inverse Gamma model can be the most used to handle linear regressions with autoregressive processes (Greenberg, 2008). Its popularity can obey to its mathematical tractability given that the conditional posterior distributions of  $\beta' = [\beta_1, \beta_2, \beta_3]$  and  $\sigma_\epsilon^2$  have closed forms, and as a consequence the Gibbs sampler can be used to simulate them.

However, this model can have two pitfalls. First, if the likelihood function is quite flat or the prior distribution is concentrated on the tails of the likelihood, using a Normal prior cannot be a good idea due to its thin-tailed property. Therefore, the posterior outcomes may be too sensitive to hyperparameters of the Normal prior (Berger, 1985). Second, the assumption that  $\sigma_e^2$  follows an Inverse-Gamma distribution can be questionable. In particular, it is commonly considered as a “non-informative” improper prior distribution for the variance parameter,  $\mathcal{IG}(e, e)$ , when  $e \rightarrow 0$  (Spiegelhalter et al., 2003). However, this prior distribution does not have any proper limiting posterior distribution. As a consequence, posterior inference is sensitive to the choice of  $e$  (Gelman, 2006). We use in our application a  $\mathcal{IG}(\alpha_0/2, \delta_0/2)$ , such that  $\alpha_0/2 = \delta_0/2 = 0.001$ , which is a common choice (Spiegelhalter et al., 2003).

Given the previous limitations of the Normal-Inverse Gamma model, and the fact that robust Bayesian analysis, which has an excellent mathematical foundation in Walley (1991), can be based on flat-tailed priors (Berger, 1985), we introduce three additional prior specifications. The Normal-Scaled Beta two and the Student’s t-Scaled Beta two, models that use as prior distributions of the variance parameter a Scaled Beta two distribution (compound Gamma distribution (Satya, 1970)). This distribution emerges when the scale parameter has a Gamma distribution which in turn is mixed through a Gamma distribution. The Scaled Beta two prior has some advantages such as its flexibility, some hyperparameters can generate heavy tails, simulation from it is easy, and it can be inside a Gibbs sampling in some circumstances. In addition, this prior distribution discounts its influence when there is conflict between prior and sample information, and leads to strong shrinkage when there is not conflict (Fúquene et al., 2014; Pérez et al., 2014). We use a  $\mathcal{SB2}(\alpha_0, \delta_0, q)$ ,  $\alpha_0 = \delta_0 = 1$  and  $q = 10$ , obtaining a bounded at the origin, heavy tail and vague prior distribution. In addition, the Student’s t-Inverse Gamma and the Student’s t-Scaled Beta two use as prior distribution for the location parameters a Student’s t distribution. It is well known that this distribution has heavier tails than the Normal distribution when there are few degrees of freedom. Specifically, we use an independent multivariate Student’s t with only six degrees of freedom ( $v = 6$ ), that is, two for each location parameter. Thus, this prior is a flat-tailed distribution which might emerge in the context of a hierarchical model whose first stage is based on a natural conjugate family where there is a scale mixture of Normal distributions.

### 3 Elicitation: the hyperparameters of the prior distributions for location parameters

We should have in mind that our point of departure are some families of prior distributions, where the priors of the variance parameter are “non-informative”. So, we follow a structural elicitation procedure (Kadane and Wolfson, 1998) to elicit the income and price demand elasticities, and the semi elasticity associated with the average number of people living in the household. The reason is that these parameters are more approachable by the expert’s knowledge.

We elicit an expert who worked for two years in the most important public utility company in the Metropolitan Area of Medellín (Colombia). In the last years, this expert has worked as consultant of this company in several projects related to estimation and forecasting of

utility demand. In addition, this expert has a degree in Economics, and two Masters degrees, Economics and Finance, and a PhD in Statistics. Finally, he has published papers associated with estimation of demand functions in the utility sector. So, we guess that this person is an expert in the piped water service with good foundations on statistics.

Regarding the elicitation procedure, the main objective is to convert the expert's knowledge into probabilistic statements: a mean elasticity or semi elasticity, and their variances. The fundamental steps in this process are (Kadane and Wolfson, 1998):

1. Establishing the general framework of the elicitation process.
2. Obtaining some characteristics of the probability distribution function of elicited parameters.
3. Checking the consistency of the expert's statements.

An important issue in an elicitation process is how people perceive reality, and the way that people assign statistical statements to events. In particular, people use heuristics to make statistical statements, and these heuristics can cause bias (Tversky and Kahneman, 1974, 1973). Obviously, these heuristics are based on available information, where recent events have a more important impact than past events. Fischhoff and Beyth (1975) have shown that prior knowledge of an event causes some distortions in the memory that can affect the elicitation procedure. Furthermore, people make estimates by starting from an initial value that is adjusted to yield a final answer. Generally, this adjustment is typically insufficient. This phenomenon is reinforced by conservatism, which means that the updating process of prior statistical statements, given new information, is too close to prior statements compared to the revision indicated by the Bayes' theorem. Moreover, Tversky and Kahneman (1971) have shown that individuals incorrectly think that the characteristics of any sample are the same as the characteristics of the population, even in the case of small samples. As we can see, the elicitation procedure has a lot of shortcomings; we try to take into account some of these in our elicitation process.

Although our expert has experience regarding piped water demand, we showed him some descriptive statistics of our main data (1985q1–2009q4). These are in Table 1 where we can see that the average monthly consumption of piped water of a household of strata four is  $20.96m^3$  and its average annual growth rate is -3.01%. The average monthly real per capita income is US\$ 437.01 and the average piped water real price is  $US\$/m^3$  0.23, both using as base month December 2000, their average annual growth rate are 1.50% and 3.44%, respectively. In addition, the average number of people in the household is 4.07 with a standard deviation equal to 0.44, and an average annual growth rate equal to -1.50%.

**Table 1** *Descriptive statistics: piped water demand of strata four in Medellín (Colombia)\**

Variable	Consumption ( $m^3$ )	Income (US\$)	Household size (people)	Water price (US\$/ $m^3$ )
Mean	20.96 (5.57)	437.01 (53.66)	4.07 (0.44)	0.23 (0.10)
Annual Growth Rate	-3.01% (3.45%)	1.50% (12.85%)	-1.50% (2.27%)	3.44% (10.65%)

\* Standard deviation in parenthesis

We introduced to our expert some basic concepts of our model, and the main objective of this research. In addition, we warned the expert about the heuristic biases, *availability*, *anchoring*, *conservatism* and *representativeness*, and gave him some training about consistency in elicited statements to mitigate the problems associated with the elicitation technique. Regarding this last point, we first implement an elicitation procedure based on the Cumulative Distribution Function, assessment of fractiles, and then, we check the consistency of the expert's statements through bets (Winkler, 1967). After this stage, we perform some feedback with the expert, and finally, we arrive to some expert's concluding statements.

To mention an example of our elicitation process, we show some part of the interview (Winkler, 1967): “Let us consider the income elasticity of piped water demand of the representative household in stratum four in Medellín (Colombia). What is the minimum ( $\beta_1^{Min}$ ) and maximum ( $\beta_1^{Max}$ ) income elasticities that you can settle for this representative household in this city? Are you sure about these limits? Are you ready to bet any quantity of money regarding the income elasticity? Are you 100% sure that you gain this bet if you select this interval?

Now, can you select a point in the interval  $[\beta_1^{Min}, \beta_1^{Max}]$  such that it is equally likely that the elasticity is less than or greater than this point ( $\beta_1^{0.5}$ )? Given this last value, can you determine a point between  $\beta_1^{Min}$  and  $\beta_1^{0.5}$  such that it is equally likely that the elasticity is less than or greater than this new point ( $\beta_1^{0.25}$ )? In addition, can you determine a value between  $\beta_1^{0.5}$  and  $\beta_1^{Max}$  such that it is equally likely that the elasticity is less than or greater than this new point ( $\beta_1^{0.75}$ )?

Select points in the interval  $[\beta_1^{Min}, \beta_1^{Max}]$  such that there are probabilities of 0.1, 0.05 and 0.01 that the price elasticity is less than these points. Now select points in the interval  $[\beta_1^{Min}, \beta_1^{Max}]$  such that there are probabilities of 0.9, 0.95 and 0.99 that the price elasticity is less than these points.”

Then, we check coherence of the expert's statements. For instance, the expert established 0.4 as the 0.25 fractile of the income elasticity, which corresponds to 1-to-3 odds ( $1/(1+3) = 0.25$ , UK format, that is fractional odds), then we settled the following betting situation: “Given that  $\beta_1$  is the actual income elasticity, there are two bets where you have to choose one:

- Bet I
  - If  $\beta_1 < 0.4$ , you win US\$2.
  - If  $\beta_1 > 0.4$ , you lose US\$1.

- Bet II
  - If  $\beta_1 > 0.4$ , you win US\$1.
  - If  $\beta_1 < 0.4$ , you lose US\$2.

So, what is your choice?"

Given that the expert established 0.4 as the 0.25 fractile of the income elasticity, the fair situation is to receive US\$4 (US\$1 (initial stake) + US\$3 (winnings)) if  $\beta_1 < 0.4$  and to lose US\$1 (initial stake) in the case that  $\beta_1 > 0.4$ . In the first option the implicit probability associated with  $\beta_1 < 0.4$  is higher than expert's beliefs (0.33 vs 0.25), so it is not a good choice because the winnings are too low, whereas in the second betting situation the implicit probability associated with  $\beta_1 > 0.4$  is lower than expert's beliefs (0.66 vs 0.75), so winnings are high, and as a consequence is a good choice. Therefore, if the expert is coherent must choose the second bet. If the choice is incoherent with the elasticity assessment, we show to the expert this incoherence, and resolve it. We proceed in this way until expert's statements were consistent.

Experts may violate the axioms of the subjective expected utility (Ellsberg, 1962; Millner et al., 2013), which is the most satisfactory ontology of subjective probability (Savage, 1954), so we perform the strategy proposed by Millner et al. (2013) to check that our expert follows these axioms.

We use  $\beta_{(l),0} = \sum_i \beta^{(i)} (F(\beta^{(i)}) - F(\beta^{(i-1)}))$  and  $B_{(l,l),0} = \sum_i (\beta^{(i)} - \bar{\beta})^2 (F(\beta^{(i)}) - F(\beta^{(i-1)}))$ ,  $l = \{1, 2, 3\}$  and  $i = \{Min, 0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99, Max\}$  to calculate the mean and variance from elicited fractiles, then we asked to the expert about his belief of  $\beta_{(l),0}$  as measure of central tendency and  $B_{(l,l),0}$  as a measure of dispersion.

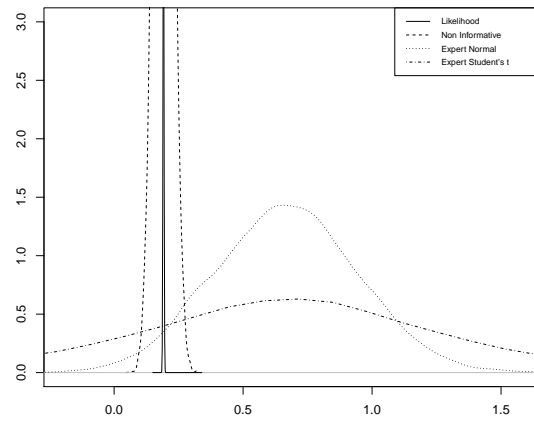
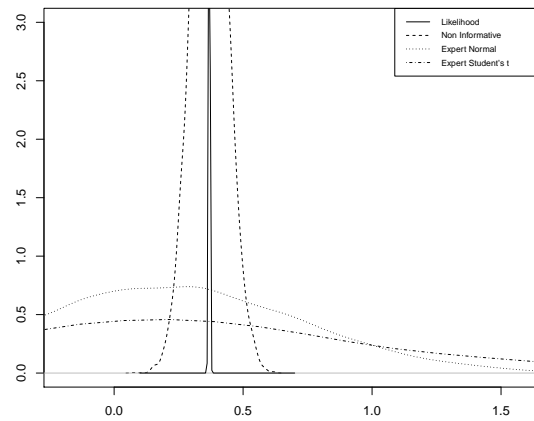
Mean and standard deviation of the elicited parameters can be seen in Table 2. In addition, we observe in this table the mean and standard deviation of the posterior distributions using non-informative priors (Chib, 1993), which imply a posterior distribution that reflects only sample information (Judge et al., 1985). For instance, the elicited mean of the price demand elasticity is equal to -0.51, whereas we obtain -0.22 using sample information. The former value means that according to the expert's information, an increase of 10% in the price implies a reduction of 4.0%  $((Exp(-0.51) - 1) \times 10)$  in water consumption. On the other hand, this price increase implies a reduction of 1.97%  $((Exp(-0.22) - 1) \times 10)$  using sample information.

Following the conventional approach of using prior distributions with well known analytical expressions, we show in Figures 1, 2 and 3 the prior distributions of the parameters using the elicited mean and standard deviation under the assumption of normal and Student's t. Additionally, we can see the posterior distributions using non-informative priors, as well as the likelihood function of each parameter conditioned to maximum of remaining parameters. As we can see the prior distributions are concentrated on the tail of the likelihood in the case of the income elasticity (Figure 1). Thus, we check robustness in an environment where there is apparently misalignment between sample information and expert's knowledge.

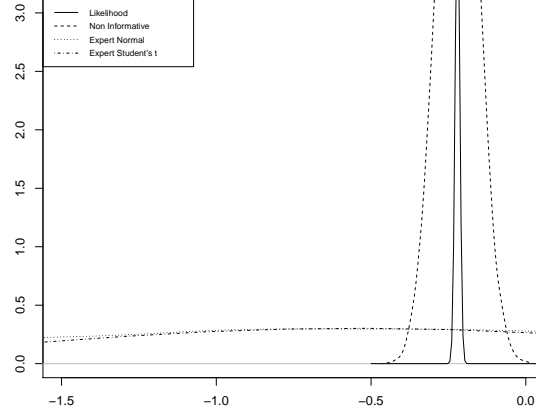
**Table 2** *Parameter estimates: Elicited and Non-informative priors.\**

Parameter	Elicitation	Non-informative Priors
Income Elasticity	0.67 (0.28)	0.19 (0.032)
Household Size Semi-Elasticity	0.18 (0.53)	0.37 (0.067)
Price Elasticity	-0.51 (1.29)	-0.22 (0.062)

\* Standard deviation in parenthesis

**Figure 1:** Distributions and Likelihood: Income elasticity**Figure 2:** Distributions and Likelihood: Household size semi elasticity



**Figure 3:** Distributions and Likelihood: Price elasticity

#### 4 Posterior distributions

The likelihood function of our model is given by

$$f(\mathbf{y} : \mathbf{x} | \boldsymbol{\beta}, \phi, \sigma_\epsilon^2) = \frac{1}{(2\pi\sigma_\epsilon^2/(1-\phi^2))^{1/2}} \text{Exp} \left\{ -\frac{(y_1 - \mathbf{x}'_1 \boldsymbol{\beta})^2}{2\sigma_\epsilon^2/(1-\phi^2)} \right\} \frac{1}{(2\pi\sigma_\epsilon^2)^{(T-1)/2}} \text{Exp} \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^T (\hat{y}_t - \hat{\mathbf{x}}'_t \boldsymbol{\beta})' (\hat{y}_t - \hat{\mathbf{x}}'_t \boldsymbol{\beta}) \right\}$$

where  $y_t = \log \{cme_t\}$ ,  $\mathbf{x}_t = [\log \{I_t\}, n_t, \log \{p_t\}]'$ ,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]'$ ,  $\hat{y}_t = y_t - \phi y_{t-1}$  and  $\hat{\mathbf{x}}_t = \mathbf{x}_t - \phi \mathbf{x}_{t-1}$ .

In addition, we assume prior independent distributions

$$\pi(\boldsymbol{\beta}, \phi, \sigma_\epsilon) = \pi(\boldsymbol{\beta})\pi(\phi)\pi(\sigma_\epsilon).$$

Therefore, the posterior distribution is given by

$$\pi(\boldsymbol{\beta}, \phi, \sigma_\epsilon | \mathbf{y}, \mathbf{x}) \propto f(\mathbf{y} : \mathbf{x} | \boldsymbol{\beta}, \phi, \sigma_\epsilon^2) \pi(\boldsymbol{\beta}) \pi(\phi) \pi(\sigma_\epsilon).$$

##### 4.1 Normal-Inverse Gamma Model

We initially assume that the prior distributions are  $\boldsymbol{\beta} \sim \mathcal{N}_K(\boldsymbol{\beta}_0, B_0)$ ,  $\sigma_\epsilon^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$  and  $\phi \sim \mathcal{N}(\phi_0, \sigma_{\phi_0}^2) I_{\phi \in (-1, 1)}$  where  $\boldsymbol{\beta}_0 = [\beta_{(1),0}, \beta_{(2),0}, \beta_{(3),0}]'$ ,  $B_0 = \text{diag} \{B_{(11),0}, B_{(22),0}, B_{(33),0}\}$  and  $I_{\phi \in (-1, 1)}$  denotes the indicator function of the set  $(-1, 1)$ . There is the second order stationary assumption on the process, which means the the mean and all covariances of  $\mu_t$  are finite and independent of time. Thus, this assumption imposes the restriction that  $\phi \in (-1, 1)$  (Chib, 1993).

It can be shown that the posterior distributions are  $\beta|y_t, x_t, \sigma_\epsilon^2, \phi \sim \mathcal{N}_K(\bar{\beta}, \bar{B})$  and  $\sigma_\epsilon^2|y_t, x_t, \beta, \phi \sim \mathcal{IG}(\alpha_1/2, \delta_1/2)$  where

$$\bar{B} = \left[ \sigma_\epsilon^{-2} \left\{ x_1 x'_1 (1 - \phi^2) + \sum_{t=2}^T \hat{x}_t \hat{x}'_t \right\} + B_0^{-1} \right]^{-1}. \quad (4.1)$$

$$\bar{\beta} = \bar{B} \left[ \sigma_\epsilon^{-2} \left\{ y_1 x_1 (1 - \phi^2) + \sum_{t=2}^T \hat{x}_t \hat{y}_t \right\} + B_0^{-1} \beta_0 \right]. \quad (4.2)$$

$$\alpha_1 = \alpha_0 + T. \quad (4.3)$$

$$\delta_1 = \delta^0 + (y_1 - x'_1 \beta)^2 (1 - \phi^2) + \sum_{t=2}^T (\hat{y}_t - \hat{x}'_t \beta)^2. \quad (4.4)$$

In addition,

$$\begin{aligned} \phi|y_t, x_t, \beta, \sigma_\epsilon^2 &\propto (1 - \phi^2)^{1/2} \text{Exp} \left\{ -\frac{1}{2\sigma_\epsilon^2} \left( (y_1 - x'_1 \beta)^2 (1 - \phi^2) \right) \right\} \\ &\times \text{Exp} \left\{ -\frac{1}{2\sigma_\epsilon^2} \left( \sum_{t=2}^T (y_t^* - \phi y_{t-1}^*)^2 \right) \right\} \\ &\times \text{Exp} \left\{ -\frac{1}{2\sigma_{\phi_0}^2} (\phi - \phi_0)^2 \right\} I_{\phi \in (-1, 1)} \end{aligned} \quad (4.5)$$

where  $y_t^* = y_t - x'_t \beta$ .

The conditional posterior distributions of  $\beta$  and  $\sigma_\epsilon^2$  can be simulated by Gibbs sampler. However, we must use a Metropolis-Hastings algorithm to draw  $\phi$ . In particular, we use as proposal density a Normal distribution whose variance and mean are the following expressions:

$$\sigma_{\phi_p}^2 = \left( \sigma_\epsilon^{-2} \sum_{t=2}^T (y_{t-1}^*)^2 + \sigma_{\phi_0}^{-2} \right)^{-1}. \quad (4.6)$$

$$\bar{\phi}_p = \sigma_{\phi_p}^2 \left( \sigma_\epsilon^{-2} \sum_{t=2}^T y_t^* y_{t-1}^* + \phi_0 \sigma_{\phi_0}^{-2} \right). \quad (4.7)$$

We retain the value if  $|\phi^c| < 1$ , and then we accept this candidate following the Metropolis-Hastings lineaments.

## 4.2 Normal-Scaled Beta two Model

In the case that  $\beta \sim \mathcal{N}_K(\beta_0, B_0)$ ,  $\sigma_\epsilon^2 \sim \mathcal{SB2}(\alpha_0, \delta_0, q)$  and  $\phi \sim \mathcal{N}(\phi_0, \sigma_{\phi_0}^2) I_{\phi \in (-1, 1)}$ . We have that  $\beta|y_t, x_t, \sigma_\epsilon^2, \phi \sim \mathcal{N}_K(\bar{\beta}, \bar{B})$  and  $\phi|y_t, x_t, \beta, \sigma_\epsilon^2$  is proportional to expression 4.5. In addition,

$$\pi(\sigma_\epsilon^2|y_t, x_t, \beta, \phi) \propto \text{Exp} \left\{ -\frac{1}{2\sigma_\epsilon^2} \left( (y_1 - x'_1 \beta)^2 (1 - \phi^2) + \sum_{t=2}^T (\hat{y}_t - \hat{x}'_t \beta)^2 \right) \right\} \frac{1}{(\sigma_\epsilon^2)^{T/2+1-\alpha_0}} \left( 1 + \frac{\sigma_\epsilon^2}{q} \right)^{-(\alpha_0 + \delta_0)}. \quad (4.8)$$

We use the same strategy than in the previous model to draw  $\beta$  and  $\phi$ . In addition, we implement a Metropolis-Hastings algorithm using as proposal density an Inverse-Gamma distribution with shape parameter  $T/2 - \alpha_0$  and scale parameter  $-\frac{1}{2} \left( (y_1 - x'_1\beta)^2(1 - \phi^2) + \sum_{t=2}^T (\hat{y}_t - \hat{x}'_t\beta)^2 \right)$ . This is due to the fact that the mode of  $\sigma_\epsilon^2$  is equal to 0.002 using a non-informative prior in our application. This implies  $\left(1 + \frac{\sigma_\epsilon^2}{q}\right)^{-(\alpha_0 + \delta_0)} \approx 1$ , then equation 4.8 is approximately proportional to an Inverse-Gamma distribution.

### 4.3 Student's t-Inverse Gamma Model

Now we assume that  $\beta \sim \mathcal{T}_K(\beta_0, B_0, v)$ ,  $\sigma_\epsilon^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$  and  $\phi \sim \mathcal{N}(\phi_0, \sigma_{\phi_0}^2)I_{\phi \in (-1,1)}$ . In this case we have that  $\sigma_\epsilon^2|y_t, x_t, \beta, \phi \sim \mathcal{IG}(\alpha_1/2, \delta_1/2)$  and  $\phi|y_t, x_t, \beta, \sigma_\epsilon^2$  is proportional to expression 4.5. Regarding the conditional distribution of  $\beta$

$$\begin{aligned} \pi(\beta|y_t, x_t, \sigma_\epsilon^2, \phi) &\propto \text{Exp} \left\{ -\frac{1}{2\sigma_\epsilon^2} \left( (y_1 - x'_1\beta)^2(1 - \phi^2) + \sum_{t=2}^T (\hat{y}_t - \hat{x}'_t\beta)^2 \right) \right\} \\ &\times \left( 1 + \frac{1}{v}(\beta - \beta_0)'B_0^{-1}(\beta - \beta_0) \right)^{-(v+K)/2}. \end{aligned} \quad (4.9)$$

We can draw  $\sigma_\epsilon^2$  from a Gibbs sampler, and  $\phi$  in the same way than the previous model. In addition, we can use a Metropolis-Hastings algorithm to draw  $\beta$ . In this case we use as proposal density a Normal distribution whose covariance matrix and mean are given by

$$\Sigma_{\beta_p} = \left( \sigma_\epsilon^{-2} \sum_{t=2}^T \hat{x}_t \hat{x}'_t + B_0^{-1} \right)^{-1}. \quad (4.10)$$

$$\bar{\beta}_p = \Sigma_{\beta_p} \left( \sigma_\epsilon^{-2} \sum_{t=2}^T \hat{x}_t \hat{y}_t + B_0^{-1} \beta_0 \right). \quad (4.11)$$

### 4.4 Student's t-Scaled Beta two Model

In this case  $\beta \sim \mathcal{T}_K(\beta_0, B_0, v)$ ,  $\sigma_\epsilon^2 \sim \mathcal{SB2}(\alpha_0, \delta_0, q)$  and  $\phi \sim \mathcal{N}(\phi_0, \sigma_{\phi_0}^2)I_{\phi \in (-1,1)}$ . Thus,  $\pi(\beta|y_t, x_t, \sigma_\epsilon^2, \phi)$ ,  $\pi(\sigma_\epsilon^2|y_t, x_t, \beta, \phi)$  and  $\pi(\phi|y_t, x_t, \beta, \sigma_\epsilon^2)$  are proportional to expressions 4.9, 4.8 and 4.5 respectively.

We use Metropolis-Hastings algorithms to draw  $\beta$ ,  $\sigma_\epsilon^2$  and  $\phi$ . Regarding  $\beta$  the proposal density is Normal with covariance matrix and mean vector given by expressions 4.10 and 4.11. In addition we use an Inverse-Gamma distribution with shape parameter  $T/2 - \alpha_0$  and scale parameter  $-\frac{1}{2} \left( (y_1 - x'_1\beta)^2(1 - \phi^2) + \sum_{t=2}^T (\hat{y}_t - \hat{x}'_t\beta)^2 \right)$  as proposal density to draw  $\sigma_\epsilon^2$ , and a normal density with variance and mean given by expressions 4.6 and 4.7 for  $\phi$ .

## 5 Application

In Table 3 and Figure 4 can be observed the results of our application. The main characteristic of this application is that the results using different models are robust, and although there

is conflict between sample information and elicited parameters, our results are similar to the results that we obtain using non-informative priors (see Table 2).

**Table 3** *Summary posterior estimates*

Normal-Inverse Gamma model				
Parameter	Mean	Median	95% Credible Interval	
			Lower	Upper
Income elasticity	0.202	0.201	0.165	0.245
Household Size semi elasticity	0.339	0.341	0.249	0.418
Price elasticity	-0.257	-0.255	-0.362	-0.163
Variance parameter	0.002	0.002	0.002	0.003
Autocorrelation coefficient	0.640	0.639	0.536	0.746
Normal-Scaled Beta two model				
Parameter	Mean	Median	95% Credible Interval	
			Lower	Upper
Income elasticity	0.202	0.188	0.164	0.244
Household Size semi elasticity	0.340	0.312	0.251	0.419
Price elasticity	-0.257	-0.289	-0.361	-0.164
Variance parameter	0.002	0.002	0.002	0.003
Autocorrelation coefficient	0.639	0.639	0.535	0.746
Student's t-Inverse Gamma model				
Parameter	Mean	Median	95% Credible Interval	
			Lower	Upper
Income elasticity	0.198	0.197	0.159	0.243
Household Size semi elasticity	0.347	0.349	0.256	0.427
Price elasticity	-0.255	-0.254	-0.359	-0.162
Variance parameter	0.002	0.002	0.002	0.003
Autocorrelation coefficient	0.636	0.636	0.530	0.743
Student's t-Scaled Beta two model				
Parameter	Mean	Median	95% Credible Interval	
			Lower	Upper
Income elasticity	0.198	0.197	0.160	0.244
Household Size semi elasticity	0.346	0.348	0.252	0.426
Price elasticity	-0.256	-0.254	-0.358	-0.163
Variance parameter	0.002	0.002	0.002	0.003
Autocorrelation coefficient	0.636	0.635	0.534	0.746

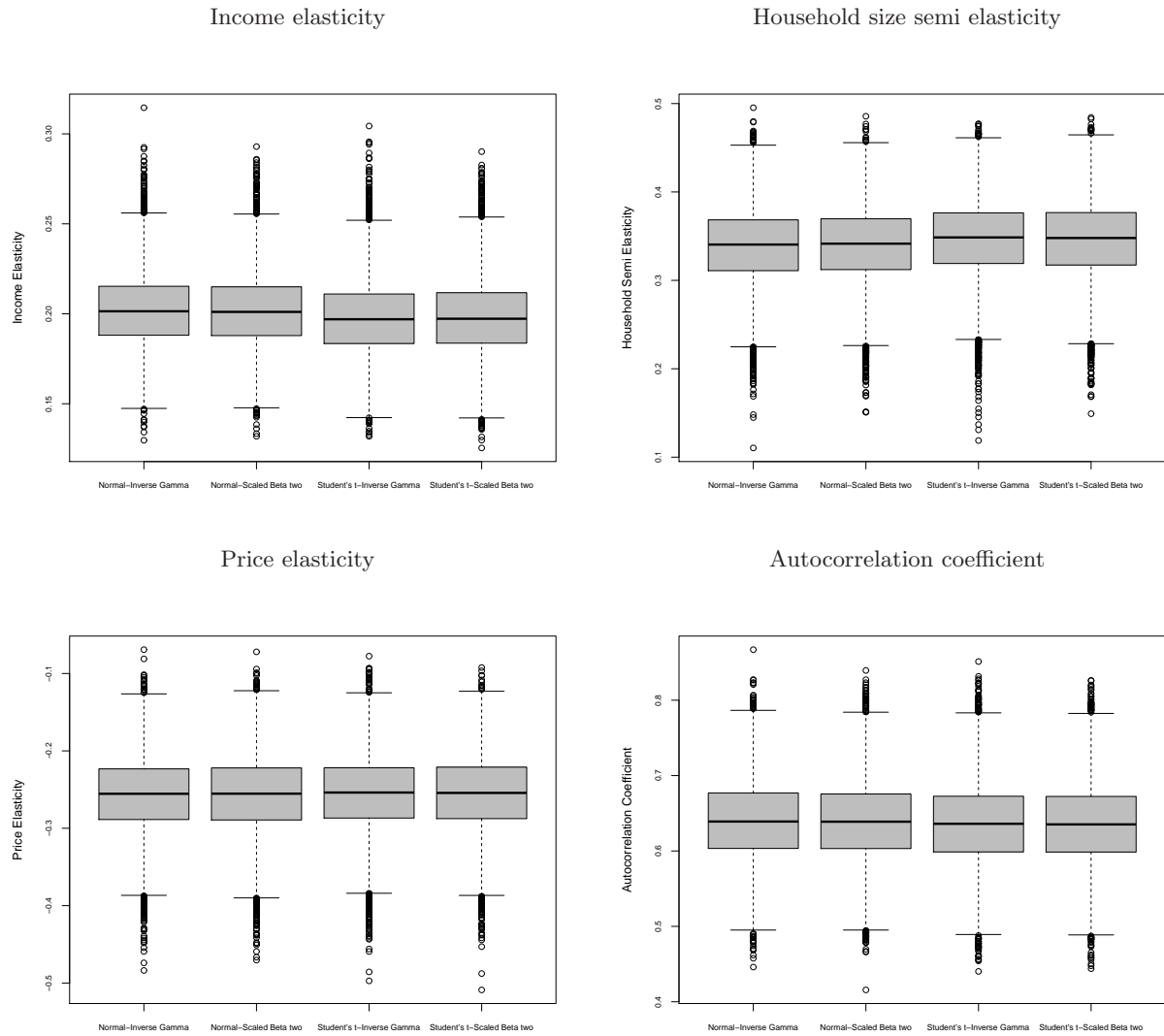
*Source:* Author's calculations

In particular, we observe that the median income elasticity is approximately equal to 0.20 with a 95% credible interval equal to (0.16, 0.24). This implies that an income increase of 1% generates an increase equal to 0.22% in piped water demand. In addition, the household size semi elasticity is equal to 0.34, this implies that one additional person in the household increases water consumption in 40.5%. Regarding the price elasticity, its mean is equal to -0.25 with a 95% credible interval equal to (-0.36, -0.16). Thus, 1% price increase implies a water consumption decrease equal to 0.22%.

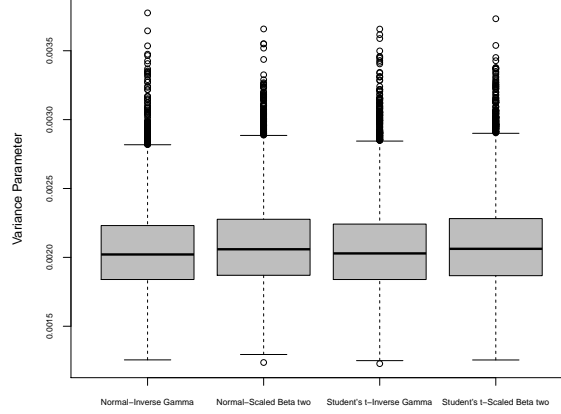
Regarding the estimation procedure, we implement the sampling algorithms using 110,000 iterations and a burn-in of 10,000. Then, we draw a sample every 10 iterations to have an effective size of 10,000. This last step is done to mitigate the autocorrelation of the chains. All the chains seem stable, and different diagnostics indicate that the chains converge to stationary distributions (see Table 4 in the Appendix Section 8. In addition, trace plots are available

upon request).

**Figure 4:** Posterior estimates: location parameters



We can see in Figure 5 the box plots associated with the posterior variance parameter. The four models are centered around 0.002, and again we see robust posterior estimates.

**Figure 5:** Posterior estimates: Variance parameter

## 6 Simulation Exercises

Although the concepts of Bayesian analysis hold valid for any sample size, it is interesting to examine the effects of the prior distributions on the posterior distributions given different sample sizes. In particular, it is well known that the prior distributions play a relatively important role when the sample size is small, although this effect tends to disappear when sample size increases (Zellner, 1996; Greenberg, 2008). Therefore, the effect of the prior distributions on Bayesian inference can be enormous when there are few data, especially when expert's knowledge is too tied around prior mean values. Under these circumstances, the method that is chosen to build the prior distributions can be very relevant.

As a consequence a possible cause that may be generating robust outcomes in our application is the joint effect of a sample size equal to 100, and a moderate level of uncertainty regarding elicited parameters. The latter fact is reflected on prior coefficients of variation equal to 0.41, 2.94 and 2.53 for the income, household size and price parameters, respectively.

To test the previous hypothesis, we perform a limited simulation exercise where we use smaller sample sizes, and assume different levels of uncertainty regarding elicited parameters. In particular, we simulate the following model

$$\log \{cme_t\} = 0.18 \log \{I_t\} + 0.38 \{n_t\} - 0.23 \log \{p_t\} + \mu_t \quad (6.1)$$

where

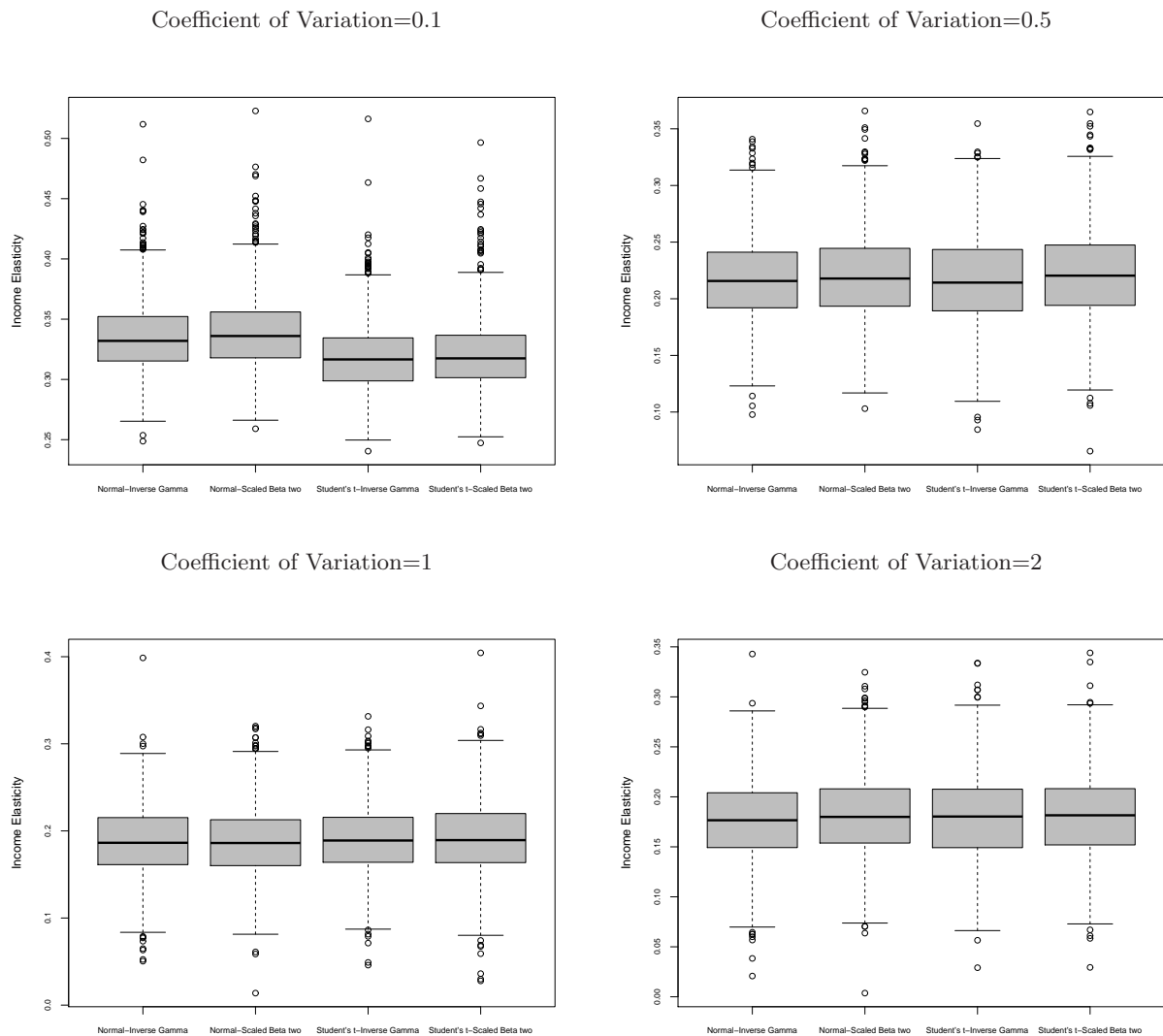
$$\mu_t = 0.61\mu_{t-1} + \epsilon_t \quad (6.2)$$

and  $t = 1, 2, \dots, T$  and  $\epsilon \stackrel{i.i.d}{\sim} \mathcal{N}(0, 0.1^2)$ ,  $\log \{I_t\} \stackrel{i.i.d}{\sim} \mathcal{N}(6.07, 0.11^2)$ ,  $n_t \stackrel{i.i.d}{\sim} \mathcal{N}(4.23, 0.53^2)$ , and  $\log \{p_t\} \stackrel{i.i.d}{\sim} \mathcal{N}(-0.46, 0.50^2)$ .

Parameters of the simulation are such that replicate data. We show the results of using a sample size equal to 25 (we try using other sample sizes, the results follow the same pattern that we show in the paper. Results available upon request).

Then we generate independent prior covariance matrices such that each one shows different degrees of uncertainty  $B_0^{(\rho)} = \text{diag} \{(|0.67|\rho)^2, (|0.18|\rho)^2, (|-0.51|\rho)^2\}$ , such that  $\rho = \{0.1, 0.5, 1, 2\}$ . Observe that 0.67, 0.18 and -0.51 are the mean elicited parameters (see Table 2).

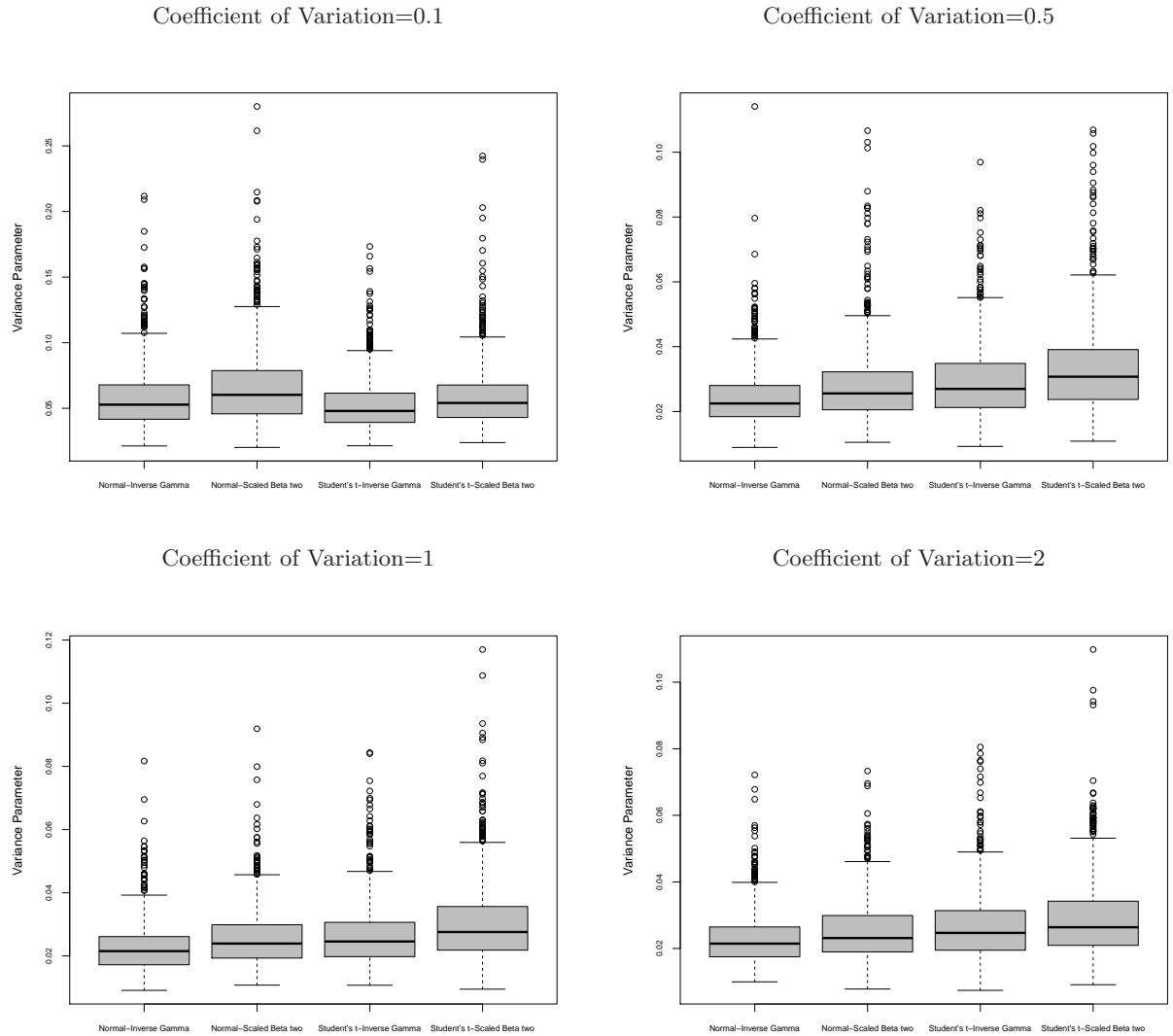
**Figure 6:** Posterior estimates: Income elasticity



We see from the box plots of the posterior income elasticity in Figure 6 that when the hypothetical expert's uncertainty is very low ( $\rho = 0.1$ ), the posterior outcomes are highly influenced by the prior mean. We observe this pattern in the four models, although is slightly

less remarkable in Student's t models. However, this pattern disappears when the level of uncertainty increases ( $\rho = \{0.5, 1, 2\}$ ). So, posterior estimates of the four model resembles the non-informative case (we observed the same pattern in the other location parameters. Available upon request).

**Figure 7:** Posterior estimates: Variance parameter



We see in Figure 7 that a high hypothetical prior expert's certainty level ( $\rho = 0.1$ ) increases model's variance. This pattern is common to the four models, although it is slightly higher in Scaled Beta two models. However, a decrease of expert's certainty level ( $\rho = \{0.5, 1, 2\}$ ) generates that posterior distributions of the variance in the four models converge to the non-informative outcome. In this case, the Student's t- Scaled Beta two model presents the highest level variability.



## 7 Concluding Remarks

We found in our application that posterior parameter distributions are robust to four prior specifications, namely Normal-Inverse Gamma, Normal-Scaled Beta two, Student's t-Inverse Gamma and Student's-Scaled Beta two. To test the hypothesis that this outcome is the result of a moderate sample size and a relatively high level of expert's uncertainty, we perform simulation exercises using smaller sample sizes and lower levels of expert's uncertainty. We show that the general pattern stays, although Student's t models are slightly less influenced by expert's knowledge when there is a high level of prior certainty, and Scaled Beta two models allow a higher level of variability.

Regarding the application, we found that the piped water demand in Medellín (Colombia) is an normal inelastic service with income and price mean elasticities equal to 0.20 and - 0.25, respectively. In addition, this service is highly affected by household size, its mean semi elasticity is 0.34.

## 8 Appendix

**Table 4** *Stationarity and Convergence diagnostics: Application*

Normal-Inverse Gamma model				
Parameter	Heidelberger (1st Part/p-value) <sup>a</sup>	Heidelberger (2nd Part) <sup>b</sup>	Geweke <sup>c</sup>	Raftery <sup>d</sup>
Income elasticity	0.384	0.000	-1.013	0.999
Household Size semi elasticity	0.443	0.001	0.538	1.020
Price elasticity	0.757	0.001	-0.431	1.010
Variance parameter	0.402	5.900E-06	1.552	0.984
Autocorrelation coefficient	0.322	0.001	-0.332	1.010
Normal-Scaled Beta two model				
Parameter	Heidelberger (1st Part/p-value) <sup>a</sup>	Heidelberger (2nd Part) <sup>b</sup>	Geweke <sup>c</sup>	Raftery <sup>d</sup>
Income elasticity	0.624	0.000402	-0.3196	0.997
Household Size semi elasticity	0.734	0.000849	0.1559	1.01
Price elasticity	0.929	0.001018	-0.1075	0.988
Variance parameter	0.872	6.04E-06	-0.4748	1.01
Autocorrelation coefficient	0.517	0.00104	-0.4928	0.997
Student's t-Inverse Gamma model				
Parameter	Heidelberger (1st Part/p-value) <sup>a</sup>	Heidelberger (2nd Part) <sup>b</sup>	Geweke <sup>c</sup>	Raftery <sup>d</sup>
Income elasticity	0.626	0.000414	0.4554	0.989
Household Size semi elasticity	0.66	0.000861	-0.1973	0.993
Price elasticity	0.907	0.000971	0.3352	0.997
Variance parameter	0.712	6.03E-06	-0.6826	1.000
Autocorrelation coefficient	0.475	0.00107	0.4453	1.000
Student's t-Scaled Beta two model				
Parameter	Heidelberger (1st Part/p-value) <sup>a</sup>	Heidelberger (2nd Part) <sup>b</sup>	Geweke <sup>c</sup>	Raftery <sup>d</sup>
Income elasticity	0.687	0.000416	-0.03641	0.971
Household Size semi elasticity	0.822	0.000867	-0.17993	0.990
Price elasticity	0.514	0.00098	-0.62794	1.030
Variance parameter	0.139	6.09E-06	0.2779	1.000
Autocorrelation coefficient	0.704	0.00107	1.052	1.010

Notes: <sup>a</sup> Null hypothesis is stationarity of the chain, <sup>b</sup> Half-width to mean ratio (threshold of 0.1), <sup>c</sup> Mean difference test z-score, <sup>d</sup> Dependence factor (threshold of 5)

## References

- Beach, L. and Swenson, R. (1966). Intuitive estimation of means. *Psychonomic Science*, 5:161–162.
- Berger, J. (2006). The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385–402.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.
- Berger, J. O. (1994). An overview of robust Bayesian analysis. *Test*, 3(1):5–124.
- Bian, G. and Tiku, M. L. (1997). Bayesian inference based on robust priors and MML estimators: Part I, symmetric location–scale distributions. *Statistics: A Journal of Theoretical and Applied Statistics*, 29(4):317–345.
- Chib, S. (1993). Bayes regression with autoregressive errors. *Journal of Econometrics*.
- Ellsberg, D. (1962). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75(4):643–669.
- Fischhoff, B. and Beyth, R. (1975). I knew it would happen: Remembered probabilities of once–future things. *Organizational Behavior and Human Performance*, 13:1–16.
- Fúquene, J., Cook, J., and Pericchi, L. (2009). A case for robust bayesian priors with applications to clinical trials. *Bayesian Analysis*, 4(4):817–846.
- Fúquene, J., Pérez, M., and Pericchi, L. (2014). An alternative to the inverted gamma for the variances to modelling outliers and structural breaks in dynamic models. *Brazilian Journal of Probability and Statistics*, 28(2):288–299.
- Garthwaite, P., Kadane, J., and O’Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of American Statistical Association*, 100(470):680–701.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534.
- Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge, first edition.
- Judge, G., Griffiths, W., Hill, C., Lutkepohl, H., and Lee, T. (1985). *Theory and Practice of Econometrics*. Jhon Wiley & Sons Ltd.
- Kadane, J. and Wolfson, L. (1998). Experiences in elicitation. *The Statiscian*, 47(1):3–19.
- Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar Straus Giroux.
- Millner, A., Calel, R., Stainforth, D., and MacKerron, G. (2013). Do probabilities expert elicitation capture scientists’ uncertainty about climate change. *Climatic Change*, 116:427–436.
- Pérez, M., Pericchi, L., and Ruíz, I. (2014). The scaled beta2 distribution as a robust prior for scales, and a explicit horseshoe prior for locations. Technical report, University of Puerto Rico, Puerto Rico.
- Satya, D. (1970). Compound gamma, beta and F distributions. *Metrika*, 16(1):27–31.
- Savage, L. (1954). *The Foundations of Statistics*. Wiley.
- Spiegelhalter, D., Best, T., Gilks, W., and Lunn, D. (2003). BUGS: Bayesian inference using Gibbs sampling. Technical report, MRC Biostatistics Unit, England. [www.mrc-bsu.cam.ac.uk/bugs/](http://www.mrc-bsu.cam.ac.uk/bugs/).
- Tversky, A. and Kahneman, D. (1971). The belief in the law of small numbers. *Psychological Bulletin*, 76:105–110.
- Tversky, A. and Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5:207–232.
- Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, 185:1124–1131.
- Walley, P. (1991). *Statistical with Imprecise Probabilities*. Chapman and Hall.
- Winkler, R. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, 62(319):776–800.
- Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. Wiley.