Bayesian Estimation of the Linear Expenditure System

Jakub Boratyński*

http://orcid.org/0000-0002-4288-2596

Abstract

We argue that the Bayesian approach is appealing to parameter estimation for multi-sector economic models, in that it (i) handles problems with many parameters and little data, (ii) allows for explicit and flexible formulation of prior knowledge on parameter values, and parameter constraints, (iii) provides full account of uncertainty of the estimates, (iv) is based on a single estimation strategy for a wide variety of models. It is an alternative to the more common maximum entropy (minimum cross entropy) techniques. We specify a Bayesian model for the linear expenditure system, and estimate it based on data for Poland, using the open source Stan package. We then extend the basic formulation to allow for changes in tastes, represented by time-varying marginal budget shares.

Keywords: Bayesian Estimation, Linear Expenditure System

*University of Lodz, e-mail: jakub.boratynski@uni.lodz.pl,

1 Introduction

Advances in software and hardware have made Bayesian estimation techniques accessible to applied economists. Even though still relatively computationally intensive, it has become a feasible alternative to the more common methods, in particular maximum entropy (minimum cross entropy) techniques. The latter methods are prominent in the field of input-output analysis, or multi-sector economic modeling in general, as they enable estimation (of missing data in particular, but also parameters of economic models) with limited data informativeness. Maximum entropy estimation is based on a very general approach applicable to a range of problems. It is also relatively straightforward to program and it allows to impose constraints on parameters in the form of systems of equations (for example, balancing constraints that an estimated input-output table should satisfy).

However, maximum entropy approach has two important limitations. First, it does not offer an inherent analysis of uncertainties (errors) of the estimated parameters. Second, as pointed out by Heckelei et al. (2008), the seemingly neutral initial assumptions concerning parameters, formulated under the generalized maximum entropy (GME) approach (Golan et al., 1996), in fact implicitly lead to preferring some parameter values over the other. Put otherwise, the generalized maximum entropy approach implicitly imposes informative priors on parameters, which cannot be deduced and shown easily from the actual formulation of the estimation problem. These two issues are "naturally" addressed under the Bayesian approach. It comes at the cost of higher computational intensity, as well as the necessity to formulate a full stochastic model, with explicit probability distributions. The latter involves specifying the likelihood function (sampling distribution), as well as formulating prior knowledge (uncertainty) about model parameters in terms of probability distributions.

In this paper we demonstrate the use of the Bayesian techniques for estimation of the Linear Expenditure System, based on annual 2000-2016 data for Poland, for 12 COICOP categories. We chose LES for its simplicity, as convenient grounds to test the estimation approach, with a view to extending it to more flexible demand systems in future research.

2 Model 1: fixed marginal budget shares

Linear Expenditure System is formulated as follows:

$$c_t \cdot w^*_{it} = \gamma_i \cdot p_{it} + \delta_i \left(c_t - \sum_j \gamma_j \cdot p_{jt} \right) \tag{1}$$

where w^*_{it} is predicted (theoretical) budget share of good i in total consumption in year t, c_t: total nominal consumption expenditure, and p_{it}: price of good i. Parameters to be estimated are:

- γ_i: "subsistence" consumption of good i
- δ_i: budget share of good i in "non-subsistence" consumption, i.e., marginal budget share, where $\sum_i \delta_i = 1$

Parameters γ_i and δ_i can be used to derive income (total expenditure) and own-price elasticities of consumption (ϵ_i and η_{it}, respectively), as:

$$\epsilon_i = \frac{\delta_{it}}{w^*_{it}} \tag{2}$$

$$\eta_{it} = \frac{1-\delta_i}{w^*_{it}} \cdot \frac{p_{it}\gamma_i}{c_t} - 1 \tag{3}$$

Note that the elasticities vary with consumption and price levels.

In empirical specifications equation 1 does not hold strictly, but includes error terms, ξ_{it}, which account for deviations of actual (observed) consumption shares, w_{it}, from theoretical ones, w^*_{it} (see Osiewalski, 2001, pp. 148-149). For example, taking multiplicative error terms one can write:

$$w_{it} = w^*_{it}(c_t, p_t; \gamma, \delta) \cdot \xi_{it} \tag{4}$$

Different variants of stochastic specification of consumption expenditure systems are discussed by Osiewalski (2001). The structure of the LES model ensures that theoretical consumption shares, w^*_{it} are between 0 and 1, as well as they sum to 1. The requirement for empirical specifications is that these properties also hold for w_{it}, i.e., after including the error terms. This requirement can be met by adopting an appropriate joint distribution of error terms.

Another general approach, which we follow in this paper, was proposed by Fry et al. (1996). Under that approach a joint probability distribution is assigned directly to observed budget shares, w_{it}, wherein that distribution, dependent on w^*_{it}, satisfies $\sum_i w_{it} = 1$ and $w_{it} \in \langle 0, 1 \rangle$. In such a case error terms are implicit, rather than explicit, as in formulation 4. Still, both approaches can be made equivalent by specific choices of probability distributions for either the error terms or observed budget shares.

We also follow Osiewalski (2001) and Fry et al. (1996) in choosing the Additive Logistic Normal (ALN) distribution, proposed by Aitchison (1982) for modeling of compositional data. ALN is imposed on the observed consumption shares, w_{it}. The ALN formulation is based on the transformation of constrained variables – the shares – to unconstrained variables, denoted v_i:

$$v_{it} = \ln \frac{w_{it}}{w_{It}}, \quad i = 1, \ldots, I-1 \tag{5}$$

where I is the number of commodities. Note that vector v_t has $I-1$ components, while vector w_t has I components. Vector v is assumed to follow a multivariate normal distribution, with vector of expected values v^*_t and variance-covariance matrix Σ, i.e.:

$$v_t \sim \mathcal{N}(v^*_t, \Sigma) \tag{6}$$

where:

$$v^*_{it} = \ln \frac{w^*_{it}}{w^*_{It}}, \qquad i = 1, \ldots, I-1 \tag{7}$$

Hence, the "sampling model" for actual budget shares, w_{it}, can be formulated as follows:

$$\begin{aligned}
w_{it} &= \frac{\exp(v_{it})}{1 + \sum_{k=1}^{I-1} \exp(v_{kt})}, \qquad i = 1, \ldots, I-1 \\
w_{It} &= \frac{1}{1 + \sum_{k=1}^{I-1} \exp(v_{kt})} \\
v_t &\sim \mathcal{N}(v^*_t, \Sigma) \\
v^*_{it} &= \ln \frac{w^*_{it}}{w^*_{It}}, \qquad i = 1, \ldots, I-1 \\
w^*_{it} &= \frac{1}{c_t} \cdot \left(\gamma_i \cdot p_{it} + \delta_i \left(c_t - \sum_j \gamma_j \cdot p_{jt} \right) \right)
\end{aligned} \tag{8}$$

The first two lines of the above formulation are an inverse of the transformation represented by equation 5. The above formulation defines the likelihood function, $\mathcal{L}(\gamma, \delta; c, p)$, which could be derived by substituting relevant terms into the probability density function for the sequence of vectors v_t. It is implicitly assumed that vectors v_t are independent for $t = 1, \ldots, T$, and hence the joint probability density for the sequence of vectors v_t is obtained by multiplying the relevant probability density functions of individual v_t vectors.

In order to obtain a full Bayesian model, one more step is needed, namely the specification of prior distributions for model parameters, including γ, δ and Σ. Prior distributions represent initial knowledge (or ignorance) regarding parameter values. To formulate prior distribution for the covariance matrix, we decompose it to $\text{diag}(\sigma) \cdot \Omega \cdot \text{diag}(\sigma)$, where σ is a $I-1$ vector of standard deviations, and Ω is a $(I-1) \times (I-1)$ correlation matrix.

The following prior distributions were selected:

$$\begin{aligned}
\gamma_i &\sim Uniform\,(0, \bar{\gamma}_i) \\
\delta &\sim Dirichlet\,([1, \ldots, 1]) \\
\sigma &\sim Half\text{-}Cauchy(0, 1) \\
\Omega &\sim LKJ(20)
\end{aligned} \tag{9}$$

The upper bound on subsistence consumption, $\bar{\gamma}_i$, is determined as the lowest observed constant-price consumption of commodity i. The Dirichlet distribution parametrized by a vector of ones represents a uniform distribution over a unit simplex (i.e., a vector of shares). This choice implies that we treat any possible set of marginal budget shares to be equally likely before seeing the data. The other two choices relate to the possible range and correlations of the (implicit) error terms. For standard deviations we adopt Half-Cauchy distribution, while for the correlation matrix – Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski et al., 2009). The latter distribution is parametrized by a single number – the value of 1 indicating full uniformity across the space of possible correlation matrices, with higher and higher values bringing us closer and closer to the no-correlation-at-all case. It proved infeasible to assume full agnosticism regarding error scale and correlations. Intuitively, it would prior "parameter space" too wide given the little informativeness of the data, leading to problems with simulating the posterior distribution. Still, the choices reflected in equations 9 allow for a wider range of errors than eventually found when confronted with the data. In other words, prior assumptions regarding implicit error terms remain weakly-informative (or even practically non-informative for what could be considered a reasonable range of errors). Nevertheless, we treat the specification of prior distribution of implicit error terms as an issue for consideration in further research.

Formulas 8 and 9 jointly constitute a Bayesian model of the linear expenditure system. Inference on unknown parameter values (γ, δ, σ, Ω), as well as their transforms (e.g. income and price elasticities, and the covariance matrix Σ) relies on determining the posterior distribution, proportional (up to a normalizing constant) to the likelihood multiplied by the prior density:

$$P(\gamma, \delta, \sigma, \Omega | c, p) \propto \mathcal{L}(\gamma, \delta, \sigma, \Omega; c, p) \cdot P(\gamma, \delta, \sigma, \Omega) \tag{10}$$

Posterior distribution is not derived analytically. Rather, samples from an approximation of the posterior distribution are obtained numerically, using Markov-Chain Monte-Carlo (MCMC) techniques. Those samples allow to plot posterior densities, or compute statistics, such as means, standard deviations, quantiles, etc., for different estimated parameters. This, among other things, contrasts the Bayesian approach to the maximum likelihood method, which primarily yields point estimates.

3 Model 2: time-varying marginal budget shares

One challenge in estimation of consumption systems is that consumer "tastes" might change in time. The need for changes in "tastes" is, for example, apparent in dynamic historical CGE simulations. This is potentially more of a problem for simpler systems, such as LES, characterized by relatively few parameters, making it more difficult to accommodate observed changes in consumption as endogenous responses to changes in income and prices.

We address this issue by allowing marginal budget shares to vary over time, according to a deterministic trend. Possible further extensions, such as time-varying subsistence consumption, or stochastic instead of deterministic trends, are left as a case for future research.

The full Bayesian model for the generalized LES model can be written as:

$$
\begin{aligned}
w_{it} &= \frac{\exp(v_{it})}{1 + \sum_{k=1}^{I-1} \exp(v_{kt})}, \qquad i = 1, \dots, I-1 \\
w_{It} &= \frac{1}{1 + \sum_{k=1}^{I-1} \exp(v_{kt})} \\
v_t &\sim \mathcal{N}(v_t^*, \Sigma) \\
v_{it}^* &= \ln \frac{w_{it}^*}{w_{It}^*}, \qquad i = 1, \dots, I-1 \\
w_{it}^* &= \frac{1}{c_t} \cdot \left(\gamma_i \cdot p_{it} + \delta_{it} \left(c_t - \sum_j \gamma_j \cdot p_{jt} \right) \right) \\
\delta_{it} &= \frac{\exp(\lambda_{it})}{1 + \sum_{k=1}^{I-1} \exp(\lambda_{kt})}, \qquad i = 1, \dots, I-1 \qquad t = 2, \dots, T \\
\delta_{It} &= \frac{1}{1 + \sum_{k=1}^{I-1} \exp(\lambda_{kt})} \\
\lambda_{it} &= \lambda_{i,t-1} + \tau_i \cdot (t-1), \qquad i = 1, \dots, I-1 \qquad t = 2, \dots, T \\
\lambda_{i1} &= \ln \frac{\delta_{i1}}{\delta_{I1}} \qquad i = 1, \dots, I-1 \\
\gamma_i &\sim \mathit{Uniform}\,(0, \bar{\gamma}_i) \\
\delta_1 &\sim \mathit{Dirichlet}\,([1, \dots, 1]) \\
\tau_i &\sim \mathcal{N}(0, 0.1) \qquad i = 1, \dots, I-1 \\
\sigma &\sim \mathit{Half\text{-}Cauchy}(0, 1) \\
\Omega &\sim \mathit{LKJ}(20)
\end{aligned}
\tag{11}
$$

Primarily the extension amounts to adding the t subscript to marginal budget shares, δ. Since δ vectors are constrained to unit simplex, the trends are not formulated directly for those variables. Instead, linear trends are imposed on auxiliary, unconstrained variables λ which are then transformed to δ, based on the same transformation as the one used with the additive logistic normal distribution. Prior distribution for the unknown trend coefficients assumes zero expected value, but is informative in that it assumes year-to-year changes in marginal budget shares would not exceed a few percentage points (imposed by adopting standard deviation equal to 0.1, for the normal prior distribution). Marginal budget shares for year 1, δ_1, are given Dirichlet distribution uniform on unit simplex.

The above formulation leads to non-linear trends for marginal budget shares, δ_{it}. At the same time it guarantees these varying shares will always remain within the $\langle 0, 1\rangle$ bounds and sum to one (over i). At the same time, model 2 is a generalization of model 1 – for $tau_i = 0$ the two models are equivalent.

4 Model implementation and results

The model has been implemented in Stan - a programming language for Bayesian inference with MCMC sampling (Carpenter et al., 2016). Stan is interfaced with other packages or programming languages used in data analysis, including R, Python, MATLAB and Stata. Conveniently, model code pretty much follows the notation provided in the previous sections, such as for formulation 11. First, in this way the user is not required to derive likelihood functions or joint prior densities – these are "assembled" based on the building blocks, such as explicit assignment of probability distributions to observed variables or prior distributions for individual parameters, equations representing data or parameter transformations, etc. Second, the user is isolated from the internals of MCMC sampling, i.e., the procedure of simulating the posterior distribution of root parameters and their transforms. The program generates a sample of parameter values from the approximated posterior distributions, available for processing and visualisation. Estimation results presented below were summarized using 20000 samples. The estimates are based on annual data for Poland on consumption expenditure and prices by 12 COICOP categories, for the years 2000-2016 (source: Eurostat).

Figure 1: Subsistence consumption: samples from prior and posterior distribution (model 1)

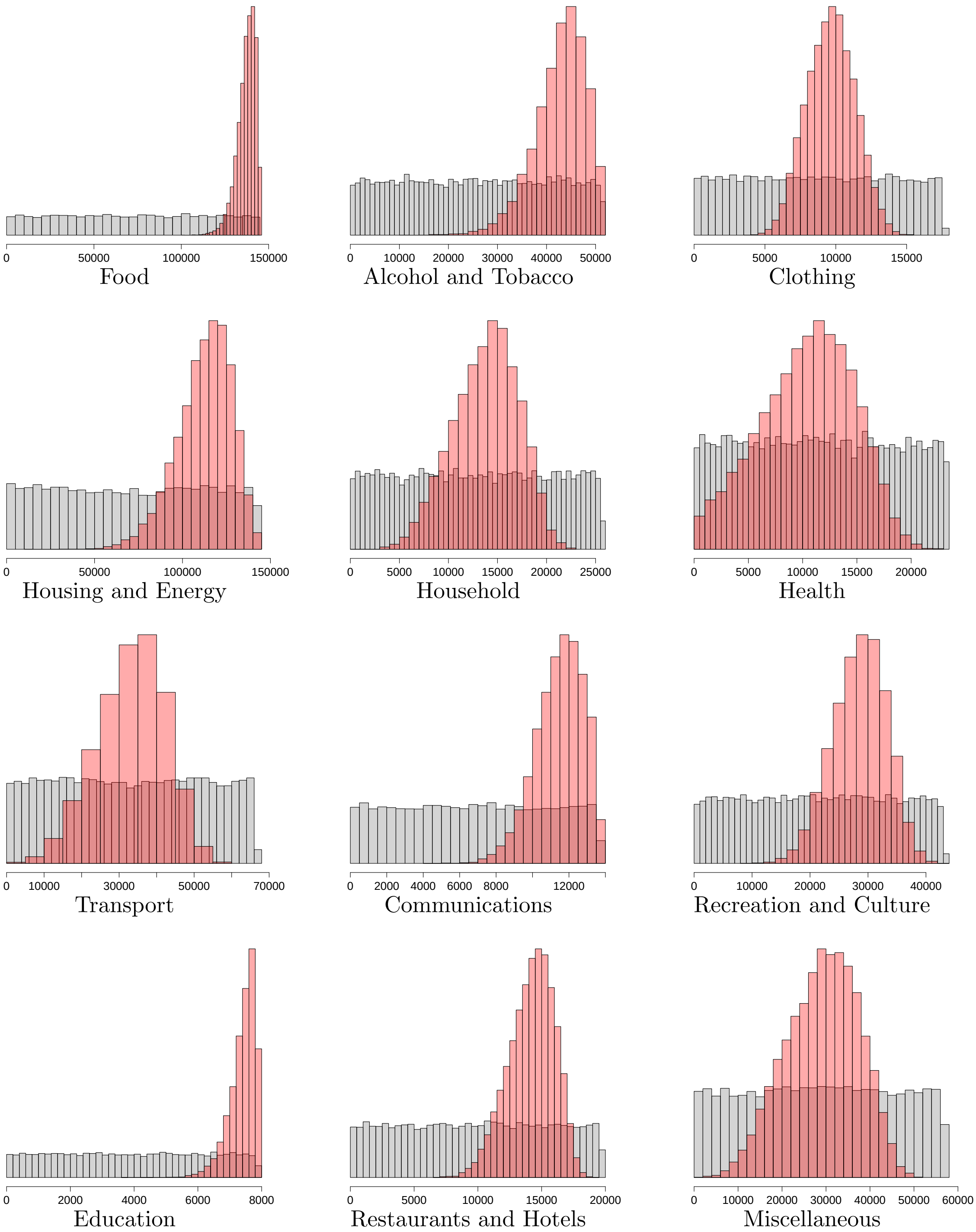


Figure 2: Marginal budget shares: samples from prior and posterior distribution (model 1)

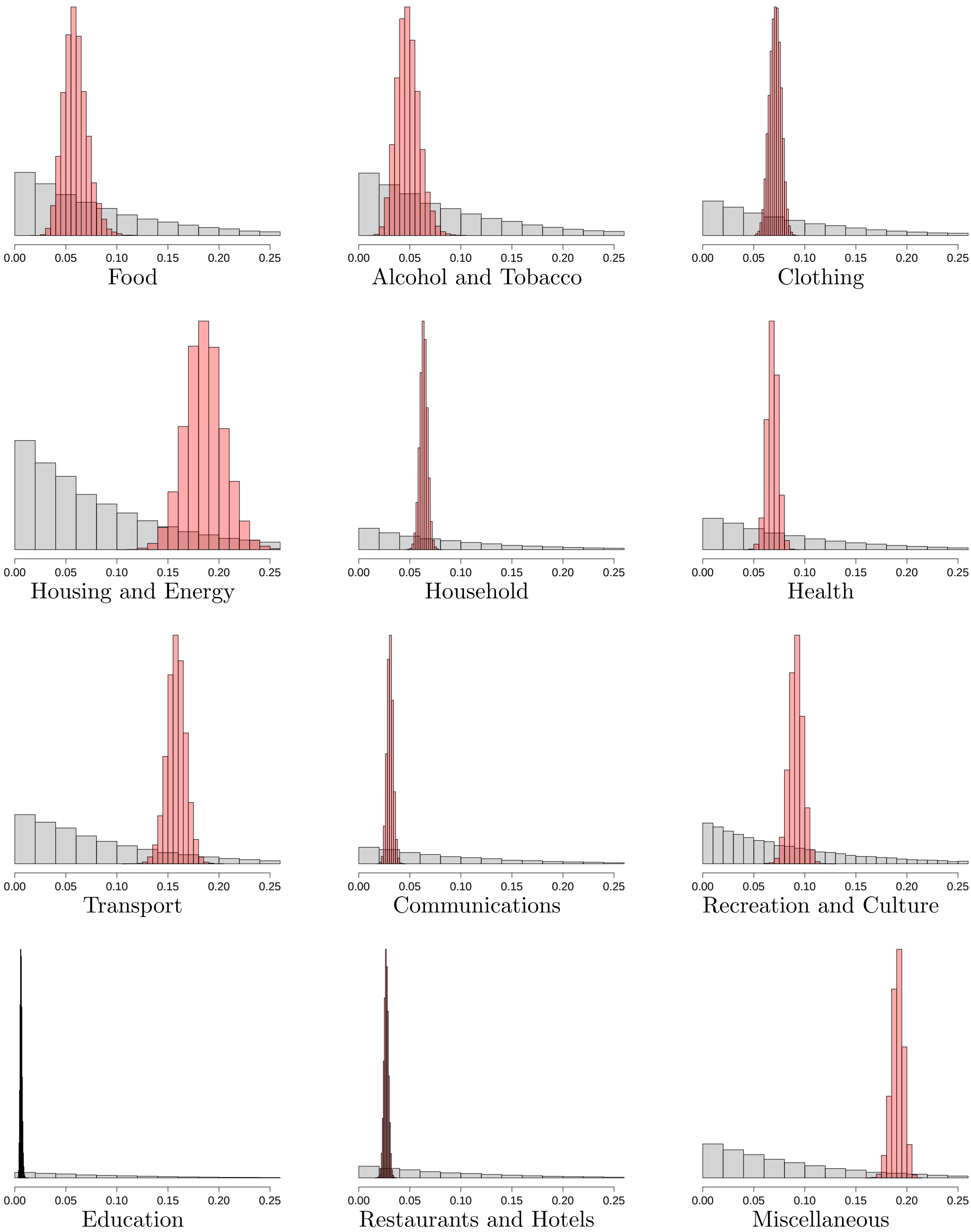


Figure 3: Income elasticities: posterior means and 5th to 95th percentile intervals (model 1)

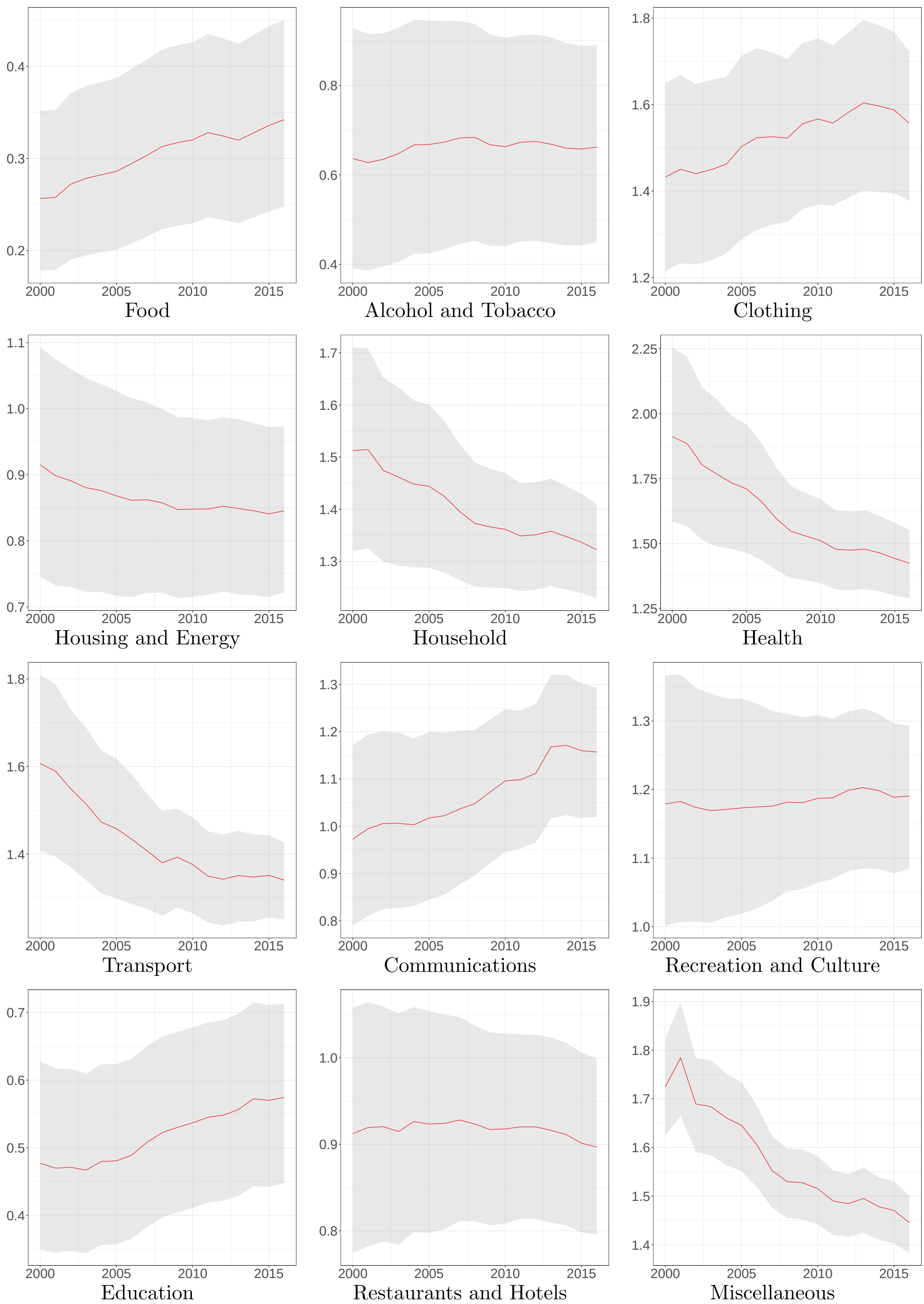


Figure 4: Price elasticities: posterior means and 5th to 95th percentile intervals (model 1)

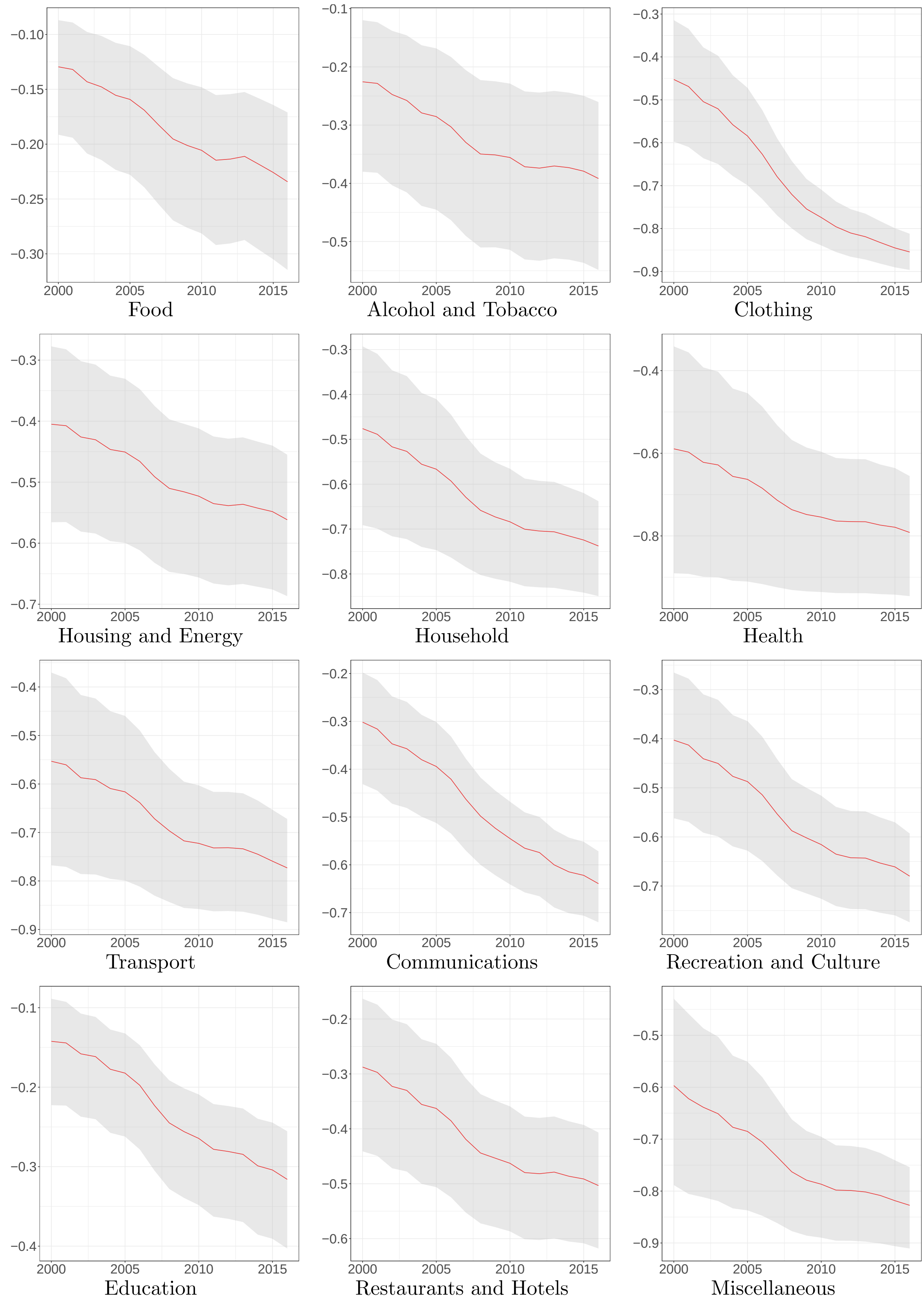


Table 1: Income and price elasticities: posterior means and standard deviations (model 1)

	Income elasticity		Price elasticity	
	Mean	St. dev.	Mean	St. dev.
Food	0.449	0.098	-0.273	0.066
Alcohol and Tobacco	0.899	0.228	-0.464	0.130
Clothing	1.638	0.105	-0.794	0.007
Housing and Energy	0.816	0.085	-0.494	0.060
Household	1.176	0.070	-0.584	0.015
Health	1.201	0.078	-0.596	0.018
Transport	1.178	0.068	-0.622	0.027
Communications	1.460	0.126	-0.706	0.089
Recreation and Culture	1.124	0.073	-0.574	0.031
Education	0.773	0.170	-0.375	0.097
Restaurants and Hotels	0.748	0.051	-0.370	0.020
Miscellaneous	1.389	0.049	-0.721	0.024

Figure 5: Subsistence consumption: samples from prior and posterior distribution (model 1 and model 2)

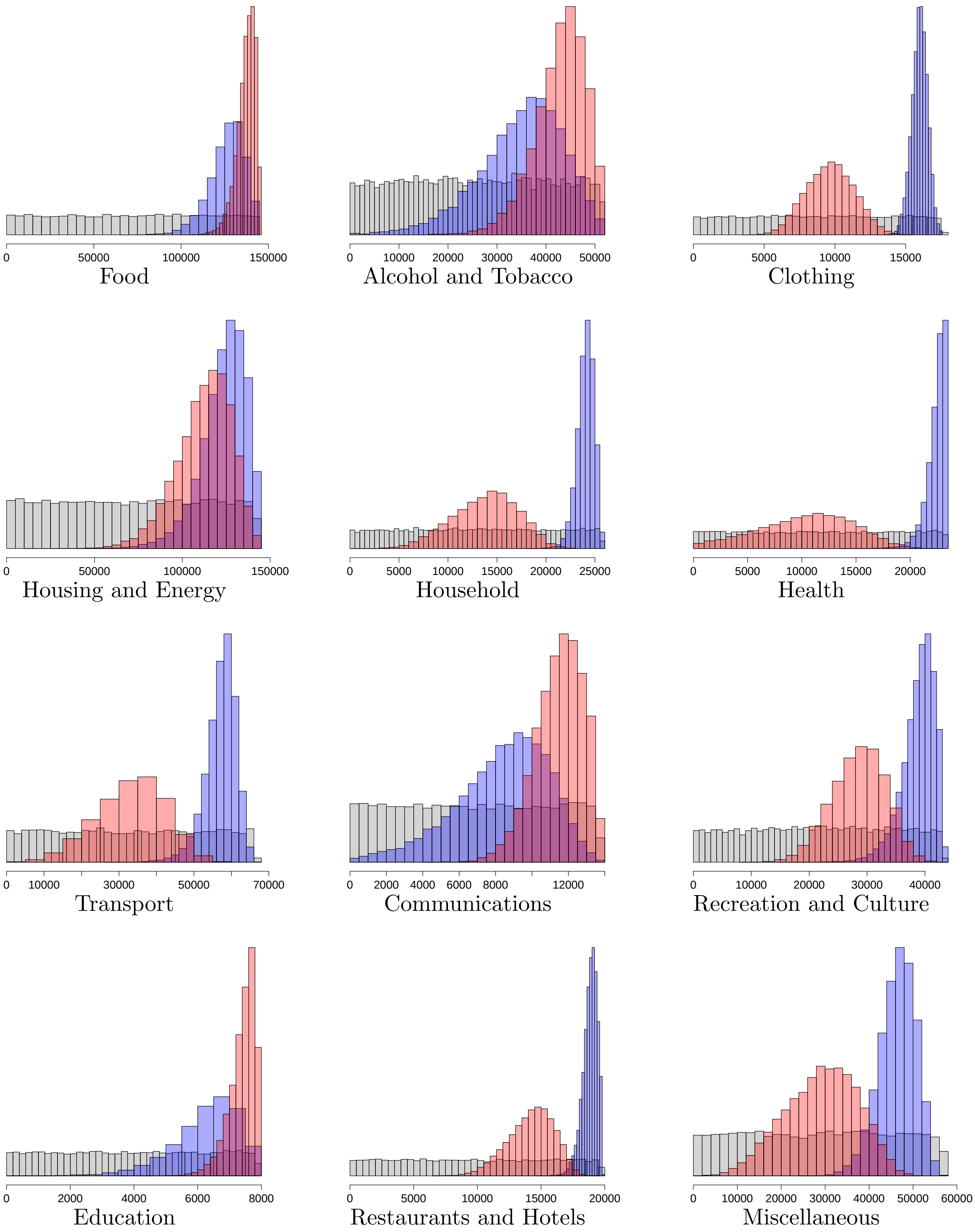


Figure 6: Marginal budget shares: posterior means and 5th to 95th percentile intervals (model 1 and model 2)

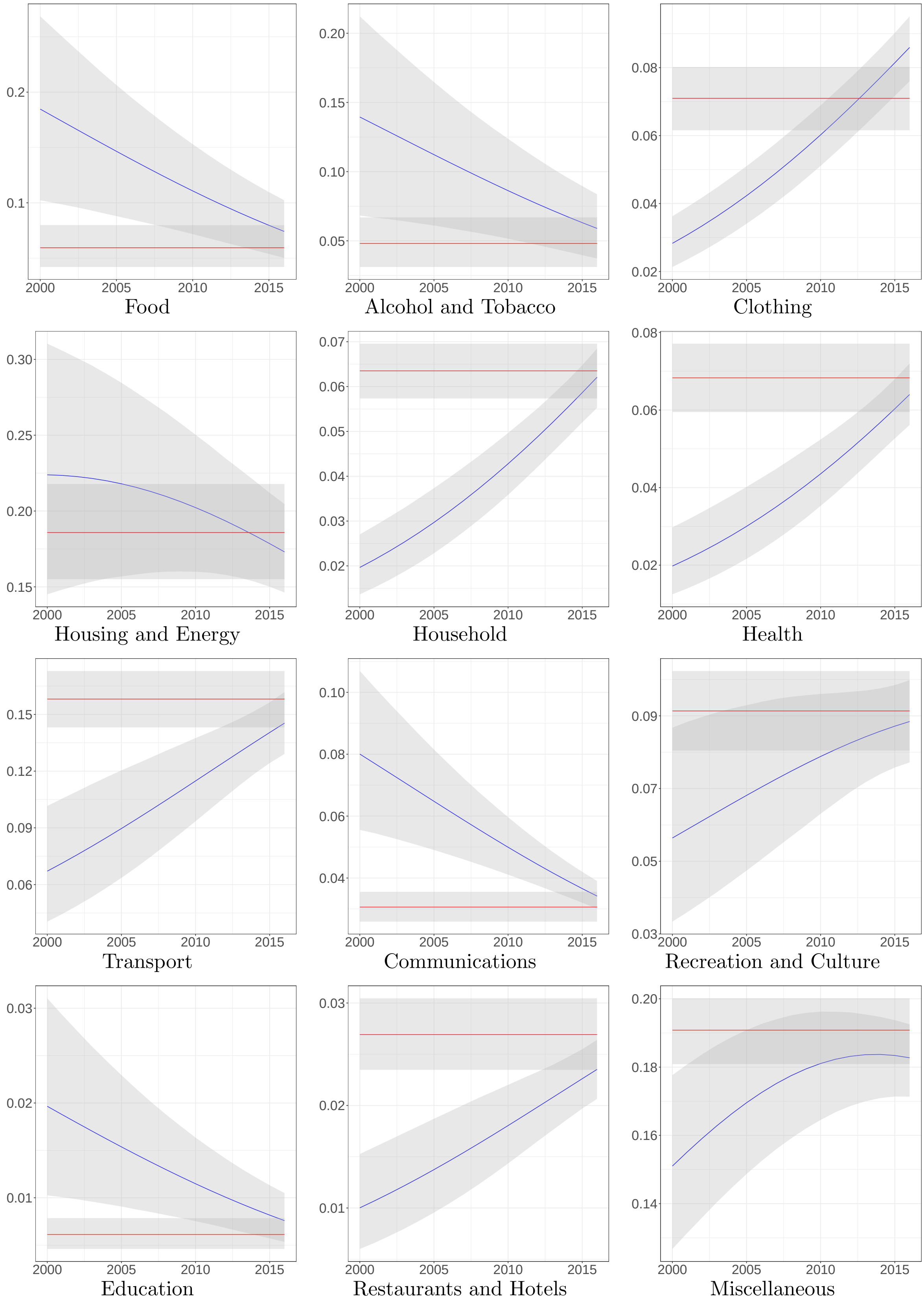


Figure 7: Income elasticities: posterior means and 5th to 95th percentile intervals (model 1 and model 2)

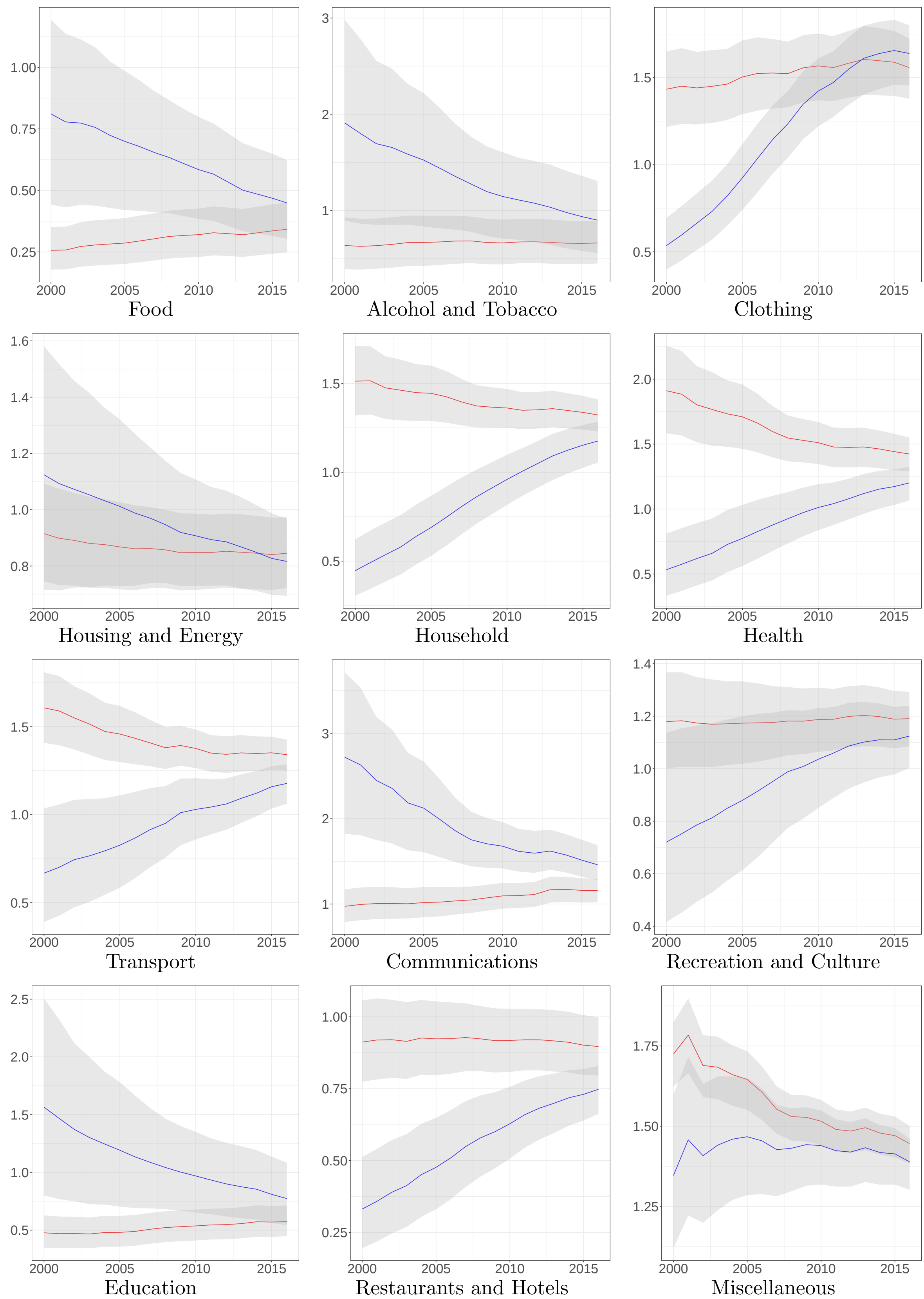


Figure 8: Price elasticities: posterior means and 5th to 95th percentile intervals (model 1 and model 2)

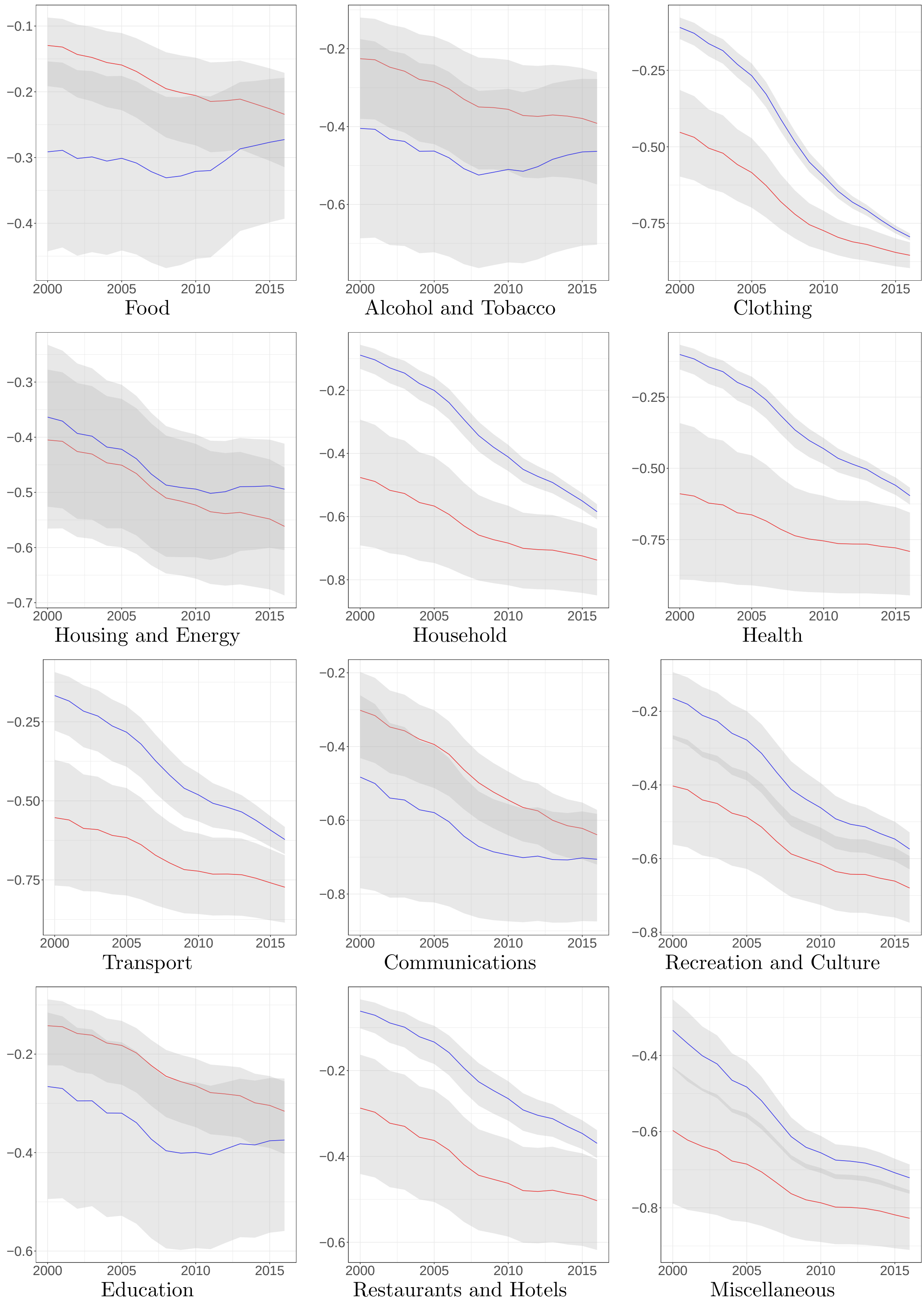
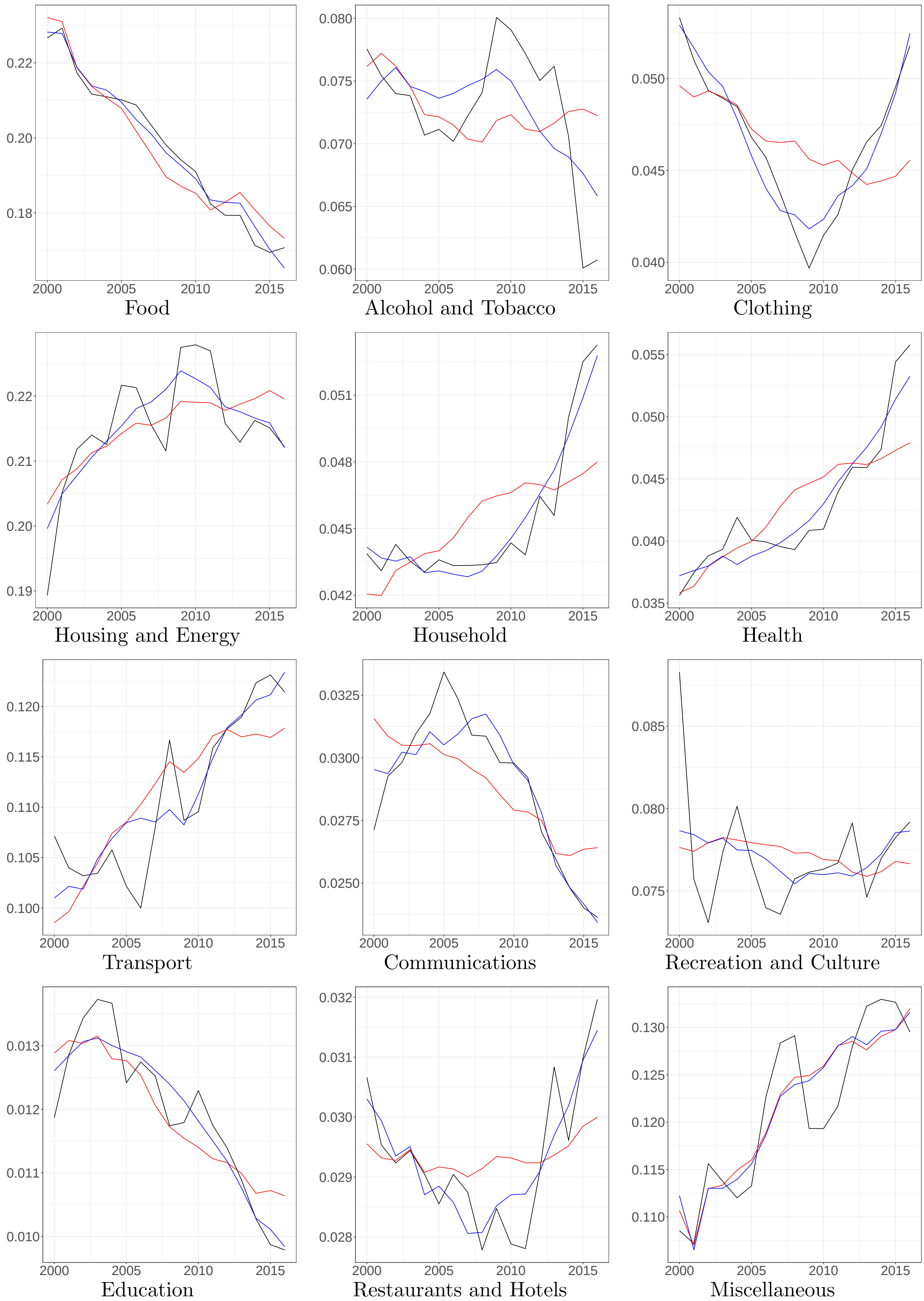


Figure 9: Models' fit: empirical budget shares and posterior mean of theoretical budget shares (model 1 and model 2)



References

John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *J Stat Softw*, 2016.

Jane M Fry, Tim RL Fry, and Keith R McLaren. The stochastic specification of demand share equations: Restricting budget shares to the unit simplex. *Journal of Econometrics*, 73(2):377–385, 1996.

Amos Golan, George Judge, and Douglas Miller. Maximum entropy econometrics: Robust estimation with limited data. 1996.

Thomas Heckelei, Ronald C. Mittelhammer, and Torbjorn Jansson. A Bayesian Alternative To Generalized Cross Entropy Solutions For Underdetermined Econometric Models. University of Bonn, Institute for Food and Resource Economics Discussion Paper No. 2, 2008.

Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001, 2009.

Jacek Osiewalski. *Ekonometria bayesowska w zastosowaniach [Applied Bayesian Econometrics]*. Wydawnictwo Akademii Ekonomicznej w Krakowie, 2001.