Marketing and Regional Sales: Evaluation of Expenditure Strategies by Spatial Sales Response Functions

Daniel Baier[1] and Wolfgang Polasek[2]

[1] Institute of Business Adminstration and Economics, BTU Cottbus, Postbox 101344, D-03013 Cottbus, daniel.baier@tu-cottbus.de
[2] IHS, Stumpergasse 56, A-1060 Vienna, polasek@ihs.ac.at

Summary. Regional marketing strategies involve usually a fixed and a flexible spending component. A recent article of Kao et al. ([1]) suggested to use a class of production functions under optimization constraints to estimate the sales effectiveness of marketing strategies. In this approach we suggest to extend this approach in two directions: First we want to model explicitly a spatial component in the production function and secondly, we will explore the use of a hierarchical model in the clustering of regional sales claims in order to optimize the geographic cost-effectiveness ratio of marketing strategies. We propose a new class of (spatial) sales-response functions based on hierarchical non-linear models that follow a CES type production function with interactions and spatial regressors for modeling the fixed (cross-sectional) and variable (time-dynamic) input components. The goal is to test the effectiveness of existing regional marketing expenditures and to suggest new expenditure patterns across time and space. The hierarchical extension of the model will be in the spirit of Rossi et al. ([2]) and models the idea, that the sales elasticities of the fixed and variable input components can vary geographically across macro-regions. The model choice in this new family of spatially enhanced sales response functions will be done in a Bayesian way using marginal likelihoods and Bayesian model averaging (BMA). The optimal clustering of the sales response function into geographic macro regions will be modeled by a normal mixture model. The modeling approach will be demonstrated using synthetic and pharma- marketing data.

Key words: Production Functions, Geographic Cost-Effectiveness, Hierarchical Spatial Models, MCMC, Model Choice by BMA.

1 Introduction

2 The multiplicative cross sectional sales response model

We will develop in the first step a cross-sectional sales response (CSSR) and then we add the spatial dimension. While the original model of Kao et al.(2005) is a panel model that estimates the responses parameter across time, we are forced to a CSSR model, because we observe only 2 time points. The CSSR model with partial derivative restrictions is defined as multiplicative model

$$y = \gamma x^{\beta_1} z^{\beta_2} e^{\epsilon} \tag{1}$$

or taking logs we find for $\beta = (\beta_0, \beta_1, \beta_2)'$ with $\beta_0 = log(\gamma)$

$$ln\ y \sim N(X\beta, \sigma_y^2 I_n) \tag{2}$$

This is a homoscedastic log-linear model with conditional mean $\mu_y = X\beta$. Adding the partial derivative restrictions for the 2 regressors, which imposes the theoretical optimality conditions that the marginal allocations should be equal across units, in a stochastic way we obtain

$$ln\ x \sim N(\mu_x, \sigma_x^2 I_n) \tag{3}$$
$$ln\ z \sim N(\mu_z, \sigma_z^2 I_n) \tag{4}$$

where the variances control the tightness of the optimality constraints: larger variances allow for more deviations from the optimal strategy. The conditional means $\mu_x = \mu_x(\beta, \lambda)$ and $\mu_z = \mu_z(\beta, \lambda)$ are given by

$$\begin{aligned} \mu_x &= (\beta_0 + ln\ \beta_1 - \lambda_1 + \beta_2 ln\ z)/(1 - \beta_1) \\ \mu_z &= (\beta_0 + ln\ \beta_2 - \lambda_2 + \beta_1 ln\ x)/(1 - \beta_2). \end{aligned} \tag{5}$$

This follows from both partial derivatives:

$$\begin{aligned} \partial y/\partial x &= y_x = \beta_0 \beta_1 x^{\beta_1 - 1} z^{\beta_2} \\ \partial y/\partial z &= y_z = \beta_0 \beta_2 x^{\beta_1} z^{\beta_2 - 1}. \end{aligned} \tag{6}$$

Since x and z are fully observed quantities (like money expenses or sales efforts via local and global advertising), these restrictions take a specific but known values for each observation, if the parameters of the SRF $(\beta, \sigma_y^2$ are fully known. Now we assume that the model can be estimated by imposing stochastic partial derivatives (SPD) constraints in the following form:

$$\begin{aligned} log(y_x) &\sim N[\lambda_1, \tau_1^2] \\ log(y_z) &\sim N[\lambda_2, \tau_2^2] \end{aligned} \tag{7}$$

Now we have 4 additional parameters and it would be interesting to estimate at least some of them - together with the SRF. The λ_i's could be interpreters

as some kind of average utility level of the sales responses while the τ_i^2's take the role of tightness parameters across observations in the sample. It seems reasonable to fix the as hyper-parameters and the average marginal utilities (AMU), i.e. the λ_i's should be estimated.
A further aspect of the SPD constraints are that by including marginal utility demons to a SRF we actually endogenize the inputs of the SRF and more complicated estimation techniques are needed.
Note that in this interpretation the SRF model is a simultaneous equation system that has cross-equation coefficients restrictions. Now the question is: By imposing the SPD constraints in the estimation of the SRF (1) would that improve the forecasting abilities of the model? Kao et al. (2005) argue that this was the case in their applications. The nest section develops a MCMC routine for the simplest case. We will report in a further paper about the Bayesian system equation approach to estimate the SRF.

2.1 Bayesian Inference by MCMC

The parameters of the model are $\theta = (\beta_0, ..., \beta_2, \lambda_1, \lambda_2, \sigma_y^2, \sigma_x^2, \sigma_z^2)$. Assuming block-wise independence, the prior distribution is given by

$$p(\theta) = \mathcal{N}[\beta \mid \beta_*, H_*] \prod_j^2 N[\lambda_j \mid \lambda_{j*}, \tau_{j*}^2] \prod_j^3 Ga[\sigma_j^2 \mid \sigma_{j*}^2 n_{j*}/2, n_{j*}/2] \tag{8}$$

We adopt the convention that all parameters with a star are known hyper-parameters of the prior distribution and those with ** are known hyper-parameters of the posterior distribution. Let $\mathcal{D} = \{y, x, z\}$ denote the observed data, then the likelihood function is

$$l(ln\ y \mid \mathcal{D}, \theta) = N[ln\ y \mid X\beta, \sigma_\epsilon^2 I_n] N[ln\ x \mid \mu_x, \sigma_x^2 I_n] N[ln\ z \mid \mu_z, \sigma_z^2 I_n] * J \tag{9}$$

where J is the appropriate Jacobian of the model. For the multiplicative model this is $J = 1 - \frac{\beta_1}{1-\beta_1}\frac{\beta_2}{1-\beta_2}$. From the posterior distribution for θ, which is proportional to the likelkihood*prior

$$p(\theta \mid \mathcal{D}) \propto l(ln\ y \mid x, z, \theta) p(\theta) \tag{10}$$

we can work out the posterior simulator for θ by MCMC.
The full conditional distributions (fcd) for the posterior are given by
1. The fcd for β

$$p(\beta \mid y, ...) \propto N[\beta \mid \beta_*, H_*] l(ln\ y \mid x, z, \theta) N[ln\ x \mid \mu_x, \sigma_x^2 I_n] N[ln\ z \mid \mu_z, \sigma_z^2 I_n] \tag{11}$$

The last 2 components contain also $\beta's$ because of the SPD constraints. Since this is not a known density we have to employ a Metropolis step, e.g. a random walk chain for the proposal β^{new}

$$\beta^{new} = \beta^{old} + N[0, c_\beta I_3] \tag{12}$$

where β^{old} is the previous generated value and c_β is a tuning constant for the variance. The acceptance probability involves the whole posterior density in (10) and is

$$\alpha(\beta^{old}, \beta^{new}) = min\left(\frac{p(\beta^{new})}{p(\beta^{old})}, 1\right), \tag{13}$$

2. The fcd for $\lambda_j, j = 1, 2$
The average utility level can be estimated in the 'usual' way.

$$p(\lambda_1 \mid y, ...) \propto N[\lambda_1 \mid \lambda_{1*}, \tau_{1*}] N[ln\ x \mid \mu_x, \sigma_x^2 I_n] \tag{14}$$

$$p(\lambda_2 \mid y, ...) \propto N[\lambda_2 \mid \lambda_{2*}, \tau_{2*}] N[ln\ z \mid \mu_z, \sigma_z^2 I_n] \tag{15}$$

where the second normal kernels can be viewed as sort of likelihood function. Again we need a Metropolis step:

$$\lambda_j^{new} = \lambda_j^{old} + N[0, c_{\lambda,j}] \tag{16}$$

where $c_{\lambda,j}$ is a small proposal variance. The acceptance probability is

$$\alpha(\lambda_j^{old}, \lambda_j^{new}) = min\left(\frac{p(\lambda_j^{new})}{p(\lambda_j^{old})}, 1\right), \tag{17}$$

where $p(.)$ is the full conditional distribution in (9).
A direct derivation shows that the pdf is a conjugate normal density:

$$\tau_{1**}^{-2} = \tau_{1*}^{-2} + \sigma_x^{-2}(1 - \beta_1)^2 \tag{18}$$

and

$$\lambda_{1**} = \tau_{1**}^2 [\tau_{1*}^{-2} \lambda_{1*} + \sigma_x^{-2}(1 - \beta_1)^2 \lambda_1]. \tag{19}$$

3. The fcd for $\sigma_j, j \in y, x, z$

$$p(\sigma_j^2 \mid y, ...) \propto Ga[\sigma_j^2 \mid \sigma_{j**}^2 n_{j**}/2, n_{j**}/2] \tag{20}$$

with

$$n_{j**} = n_{j*} + n$$

and

$$n_{j**}\sigma_{j**}^2 = n_{j*}\sigma_{j*}^2 + e_j' e_j$$

where $e_j = lnj - \mu_j$ being the current residuals of the 3 regression equations and for $j \in y, x, z$.
Finally, MCMC in the CSSR model takes the following steps:

1. Starting values: set $\beta = \beta_{OLS}$ and $\lambda = 0$

2. Draw σ_y^{-2} from $\Gamma[\sigma_y^{-2} \mid s_{y**}^2, n_{y**}]$
3. Draw σ_x^{-2} from $\Gamma[\sigma_x^{-2} \mid s_{x**}^2, n_{x**}]$
4. Draw σ_z^{-2} from $\Gamma[\sigma_z^{-2} \mid s_{z**}^2, n_{z**}]$
5. Draw λ_j from $p(\lambda_j \mid \lambda_{j**}, \sigma_{j**}^{-2})$
6. Draw β from $p\,[\beta \mid \mathbf{b}_*, \mathbf{H}_*]\,l(\theta \mid y)$
7. Repeat until convergence.

The marginal likelihood is computed by the Newton-Raftery formula

$$\hat{m}(\mathbf{y} \mid ...)^{-1} = \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} l(ln\ y \mid \mathcal{D}, \theta)^{-1} \tag{21}$$

with the parameters given for iteration i by $\theta_{(i)}$ and the likelihood in (33). Note: Extension of the model involving a market competetion variable: Let s be the vetcor of share of the sales of the own product in each region, then we can add the log share variable in the model equation:

$$ln\ y = \beta_0 + \beta_1 ln\ x + \beta_2 ln\ z + \beta_3 ln\ s + \epsilon \tag{22}$$

Conveniently, this adds one more β parameter in the estimation procedure and one more data set in d.

3 A spatial auto-regressive extension of the sales response model

Since the seminal work by Anselin (1988), spatial interactions have become an important tool in econometrics. Spatial applications have become popular in applied sciences,like in economics and also social sciences.

3.1 Spatial lags

Consider a regression model where the dependent variable $\mathbf{y} = (y_1, \ldots, y_n)'$ is not independently observed but can be spatially correlated given the $n \times K$ matrix of independent observations $\mathbf{X}$. To model the spatial dependence we have to know (or specify) a spatial weight matrix $\mathbf{W}$ which has 3 properties:

1. All entries are positive,
2. The main diagonal elements are zero, and
3. All row sums are 1, i.e.

$$\mathbf{W1}_n = \mathbf{1}_n$$

where $\mathbf{1}_n$ is a $n \times 1$ vector of ones.

Such a weight matrix could be a distance matrix if the y's are observed at geographical locations, it could the first nearest neighbor only, but also a set of all contiguous neighbors. More can be found in Anselin (1998). This allows us to specify a spatial lag variable of the dependent variable

$$\tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}. \tag{23}$$

Each element of $\tilde{\mathbf{y}}$, i.e., $\tilde{y}_j = \mathbf{w}_j\mathbf{y}$ is a new "neighborhood observation", which summarizes the influence of the neighbors in form of a weighted average of the dependent variable and the j^{th} row vector $\mathbf{w}_j$. Therefore we can formulate a 'structural' form of the spatial SAR model in the following form:

$$\mathbf{y} = \mathbf{X}\beta + \rho\mathbf{W}\mathbf{y} + \epsilon, \quad \epsilon \sim \mathcal{N}\left[\mathbf{0}, \sigma^2\mathbf{I}_n\right], \tag{24}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and ρ is the spatial correlation parameter. If ρ is zero then the model reduces to a simple regression model with independent errors. (Therefore we could test for spatial dependence by testing the restriction $\rho = 0$).
Next, we obtain a reduced form if we shift all dependent variables on the left hand side:

$$\mathbf{z} = \mathbf{y} + \rho\mathbf{W}\mathbf{y} = \mathbf{X}\beta + \epsilon, \quad \epsilon \sim \mathcal{N}\left[\mathbf{0}, \sigma^2\mathbf{I}_n\right], \tag{25}$$

Using the spread matrix $\mathbf{R}$ and its inverse

$$\mathbf{R}^{-1} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}.$$

we obtain the reduced form

$$\mathbf{y} \sim \mathcal{N}\left[\mathbf{R}^{-1}\mathbf{X}\beta, \sigma^2(\mathbf{R}'\mathbf{R})^{-1}\right], \tag{26}$$

because $Var(\mathbf{R}\epsilon) = \sigma^2\mathbf{R}\mathbf{R}'$. The prior distribution for the parameter $\theta = (\beta, \sigma^{-2}, \rho)$ is given by the product of (independent) blocks of normal and gamma distributions:

$$p(\beta, \sigma^{-2}, \rho) = p(\beta) \cdot p(\sigma^{-2}) \cdot \text{Unif}[\rho \mid -1, 1] \tag{27}$$

$$= \mathcal{N}\left[\beta \mid \mathbf{b}_*, \mathbf{H}_*\right] \cdot \Gamma[\sigma^{-2} \mid s_*^2, n_*]\frac{1}{2}, \tag{28}$$

where $Unif[-1,1]$ stands for a uniform distribution in the interval $(-1,1)$. Because of restrictions, the interval of feasible ρ's depends on λ_{min} and λ_{max}, the minimum and maximum eigenvalue of $\mathbf{W}$. It can be shown $\lambda_{min}^{-1} < 0$ and $\lambda_{max}^{-1} > 0$ and therefore ρ_t must lie between these bounds. Therefore, we restrict the prior space of ρ to the interval $(\lambda_{min}^{-1}, \lambda_{max}^{-1})$. The joint distribution for $\mathbf{y}$ and the parameter $\theta = (\beta, \sigma^2, \rho)$ is

$$p(\beta, \sigma^{-2}, \rho, \mathbf{y}) \propto \mathcal{N}\left[\mathbf{y} \mid \mathbf{X}\beta, \sigma^2\right] \cdot \mathcal{N}\left[\beta \mid \mathbf{b}_*, \mathbf{H}_*\right] \cdot \Gamma[\sigma^{-2} \mid s_*^2, n_*]. \tag{29}$$

3.2 The CSSR-SAR model

The CSSR-SAR model is the CSSR model as in (49) with a spatial lag:

$$
\begin{aligned}
&ln\ y \sim N(\mu_y = \rho W ln\ y + X\beta, \sigma_y^2 I_n) \\
&\quad\text{or} \\
&ln\ y = \rho W ln\ y + \beta_0 + \beta_1 ln\ x + \beta_2 ln\ z + \epsilon
\end{aligned} \tag{30}
$$

with $\epsilon \sim N(0, \sigma_y^2 I_n)$. The partial derivative restrictions in (3) for the 2 regressors stay the same:

$$
ln\ x \sim N(\mu_x, \sigma_x^2 I_n), \quad ln\ z \sim N(\mu_z, \sigma_z^2 I_n) \tag{31}
$$

The parameter vector now is $\theta = (\beta_0, ..., \beta_2, \lambda_1, \lambda_2, \sigma_y^2, \sigma_x^2, \sigma_z^2, \rho)$ and includes the spatial ρ. The prior is - proportionally - the same (constant) since we assume uniform prior for ρ: Unif$[\rho \mid -1, 1] = 1/2$. The reduced form of the model is

$$
ln\ y \sim \mathcal{N}\left[\mathbf{R}^{-1}\mathbf{X}\beta, \sigma^2(\mathbf{R}'\mathbf{R})^{-1}\right], \tag{32}
$$

because $Var(\mathbf{R}\epsilon) = \sigma^2\mathbf{R}\mathbf{R}'$. This expression will now be used in the likelihood function

$$
l(ln\, y \mid \mathcal{D}, \theta) = N[ln\, y \mid X\beta, \sigma_\epsilon^2 (R'R)^{-1}] N[ln\, x \mid \mu_x, \sigma_x^2 I_n] N[ln\, z \mid \mu_z, \sigma_z^2 I_n] * J \tag{33}
$$

where J is the Jacobian of the model as before. For MCMC we can use the fcd results of the previous section we just have to specify the additional fcd for ρ :
4. The fcd for ρ:

$$
p(\rho \mid y, ...) \propto |\mathbf{I}_n - \rho \mathbf{W} \mid \exp\left(-\frac{\epsilon_\rho' \epsilon_\rho}{2\sigma_y^2}\right) \tag{34}
$$

where the residuals of the spatial regression are

$$
\epsilon_\rho = ln\ y - \mathbf{X}\beta - \rho W ln\ y.
$$

We make a normal proposal:

$$
\rho^{new} = \rho^{old} + c_\rho \phi, \quad \phi \sim \mathcal{N}[0, 1]. \tag{35}
$$

The scalar c_ρ is a tuning parameter and ρ^{old} is the parameter of the previous value. The acceptance probability is

$$
\alpha(\rho^{old}, \rho^{new}) = min\left(\frac{p(\rho^{new})}{p(\rho^{old})}, 1\right), \tag{36}
$$

where p is the full conditional distribution in (34).

Finally, the MCMC procedure has just to add one more draw for the ρ parameter:

1. Starting values: set $\rho = 0, \beta = \beta_{OLS}$ and $\lambda = 0$
2. Draw σ_y^{-2} from $\Gamma[\sigma_y^{-2} \mid s^2_{y**}, n_{y**}]$
3. Draw σ_x^{-2} from $\Gamma[\sigma_x^{-2} \mid s^2_{x**}, n_{x**}]$
4. Draw σ_z^{-2} from $\Gamma[\sigma_z^{-2} \mid s^2_{z**}, n_{z**}]$
5. Draw λ_j from $p(\lambda_j \mid \lambda_{j**}, \sigma^{-2}_{j**})$
6. Draw β from $p\,[\beta \mid \mathbf{b}_*, \mathbf{H}_*]\, l(\theta \mid y)$
7. Draw ρ using $p(\rho \mid \beta, \sigma_y^{-2})$
8. Repeat until convergence.

4 Simulation of data for the SRF

In this section we show how to simulate from a bivariate sales response function (SRF2) where the x regressor is generated according to the stochastic derivative constraint (SDC): $\partial y/\partial x \sim N[\mu_x, \tau_x^2]$ where τ_x^2 is the variance of the constraint, indicating the looseness or strength of the optimality enforcement.
The bivariate sales response function (SRF2) has the following form

$$y = \gamma x^{\beta_1} e^{\epsilon} \tag{37}$$

or taking logs we find with $\beta_0 = log(\gamma)$

$$ln\ y \sim N(\mu_y = X\beta, \sigma_y^2 I_n) \tag{38}$$

Adding the partial derivative restrictions for the 2 regressors, which imposes the theoretical optimality conditions that the marginal allocations (and therefor utility) should be equal across units: in a stochastic framework we impose the condition

$$ln\ x \sim N(\mu_x, \tau_x^2 I_n) \tag{39}$$

which implies an endogeneity for x.

Example 1. : The bivariate sales response function (SRF2)
We specify the bivariate SRF by $y = e^{\beta_0} * x^{\beta_1}$ or

$$ln\ y \sim N(\mu_y = \beta_0 + \beta_1 ln\ x, \sigma_\epsilon^2 I_n) \tag{40}$$

$$\sim N(\mu_y = 2 + 1.5 ln\ x, .9^2 I_{20}) \tag{41}$$

with sample size $n = 20$, coefficients $\beta_0 = 2$ and $\beta_1 = 1.5$, residual variance $\sigma_y^2 = .81$. The derivative constraint has mean $\lambda = 3$ and variance $\tau_x^2 = 1$.

The first derivative is

$$\partial y\ /\partial x = e^{\beta_0}\beta_1 * x^{\beta_1 - 1} = \lambda \tag{42}$$

or

$$ln\ \lambda = ln\ \beta_0 + ln\ \beta_1 + (\beta_1 - 1) ln\ x \sim N[\mu_\lambda, \tau_\lambda^2].$$

The x-regressor is simulated by making x in the constraint as the dependent variable

$$\mu_x = (ln\ \beta_0 + ln\ \beta_1 - ln\ \lambda)/(\beta_1 - 1)$$

The whole regressor is simulated from a normal density:

$$ln\ x \sim N(\mu_x, \sigma_x^2 I_{20})$$

with $\sigma_x^2 = \tau_\lambda^2/(\beta_1 - 1)^2$.
From the figures 1-3 we see that the regression coefficients are far away from $\beta_0 = 2$ and $\beta_1 = 1.5$.

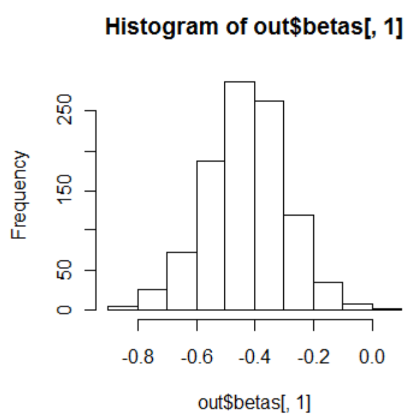
5 R routines

In the appendix we have listed the R Program:

```
#Generate a biv. constraint SR function SRF2:
#W.Polasek, March09
# see Kao et al 05 and BP09
#
#1. Parameter set up
n=20 # sample size
#target is biv. response function y = b0*x^b1 with
b0=2 #and
b1=1.5
# regressor generator is
lam = 3 #and
siglam = 1
#
#1. generate n=20 x-observations
# x-mean = (ln b0 + ln b1 - ln lam)/(b1-1)
xm = (log(b0) + log(b1) -log(lam))/(b1-1)
 #variance of x is
xvar = siglam^2/(1-b1)^2
xsig=sqrt(xvar)
#
# 2. Simulation of log x and log y
```

```
#
lx = rnorm(n,xm,xsig)
hist(lx)
yvar =.81
ysig = sqrt(yvar)
#
ym = log(b0)+ b1*lx
ym
ly = rnorm(n,ym,ysig)
#
#simple log regression
#
ll =lsfit(ly,lx)
ll$coefficients
v=var(ll$residuals)
```

Fig. 1. beta0



6 MCMC in the SRF2 model

Following the reasoning in Kao et al. (2005), see section 2, the likelihood function is

$$l(ln\ y \mid \mathcal{D}, \theta) = N[ln\ y \mid \mu_y, \sigma_\epsilon^2 I_n] N[ln\ x \mid \mu_x, \sigma_x^2 I_n] * J \tag{43}$$

with $\mu_y = \beta_0 + \beta_1 ln\ x$ and $\mu_x = (ln\ \beta_0 + ln\ \beta_1 - \mu_\lambda)/(\beta_1 - 1)$ and $\sigma_x^2 = \tau_\lambda^2/(\beta_1 - 1)^2$, since the whole regressor is simulated from a normal density:

Fig. 2. beta1

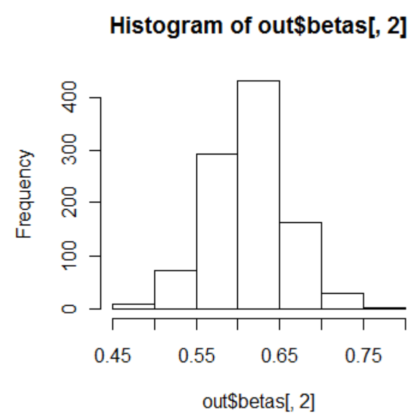
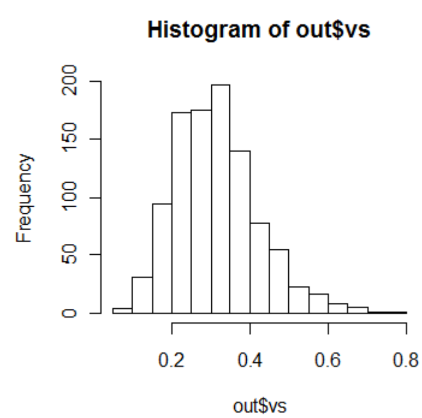


Fig. 3. variance



$ln\ x \sim N(\mu_x, \sigma_x^2 I_{20})$.

The log posterior distribution for β is

$$p(\beta \mid \mathcal{D}) \propto (ln\ y - \mu_y)^2/\sigma_\epsilon^2 + (ln\ x - \mu_x)^2/\sigma_x^2 + n * ln\ \sigma_x^2 \tag{44}$$

6.1 SRF2: MCMC for simulated data

We have generated $n = 20$ observations. We make the following assignments of hyper parameters:

The MCMC iteration is done in the following order

1. The fcd for λ The average utility level can be estimated in the 'usual' way.

$$p(\lambda \mid y, ...) \propto N[\lambda \mid \lambda_{1*}, \tau_{1*}]N[ln\ x \mid \mu_x, \sigma_x^2 I_n] \tag{45}$$

2. The fcd for β

$$p(\beta \mid y, ...) \propto N[\beta \mid \beta_*, H_*]l(ln\ y \mid x, z, \theta)N[ln\ x \mid \mu_x, \sigma_x^2 I_n] \tag{46}$$

3. The fcd for $\sigma_j, j \in y, x, z$

$$p(\sigma_j^2 \mid y, ...) \propto Ga[\sigma_j^2 \mid \sigma_{j**}^2 n_{j**}/2, n_{j**}/2] \tag{47}$$

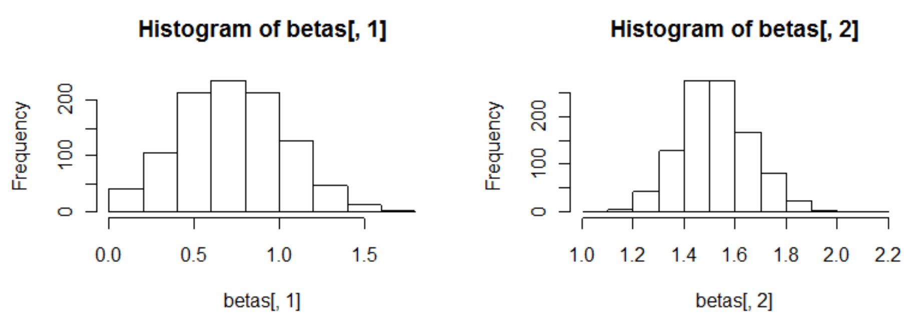
   with $n_{j**} = n_{j*} + n$ and $n_{j**}\sigma_{j**}^2 = n_{j*}\sigma_{j*}^2 + e_j'e_j$, where $e_j = lnj - \mu_j$ being the current residuals of the 3 regression equations and for $j \in y, x, z$.

Finally, MCMC in the SRF2 model takes the following steps:

1. Starting values: set $\beta = \beta_{OLS}$ and $\lambda = 0$
2. Draw λ from $p(\lambda \mid \lambda_{**}, \tau_{**}^{-2})$
3. Draw β from $p\,[\beta \mid \mathbf{b}_*, \mathbf{H}_*]\,l(\theta \mid y)$
4. Draw σ_y^{-2} from $\Gamma[\sigma_y^{-2} \mid s_{y**}^2, n_{y**}]$
5. Draw σ_x^{-2} from $\Gamma[\sigma_x^{-2} \mid s_{x**}^2, n_{x**}]$
6. Repeat until convergence.

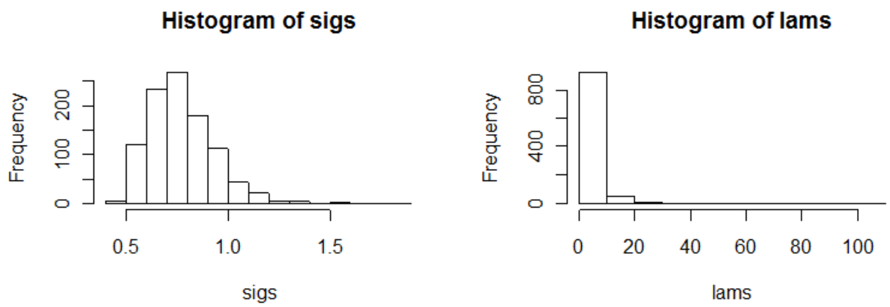
The results for 1000 repetitions are: The acceptance rate is 55.3 % The mean (and sd) of the betas are: 0.725 (.304) and 1.525 (.138)
The distribution of the parameters are

Fig. 4. beta0 and beta1



The means of the σ's and λ's across the simulation are:
mean(lams) = 3.8652
mean(sigma) = 0.7736
The main simulation program is

Fig. 5. lambda and sigma

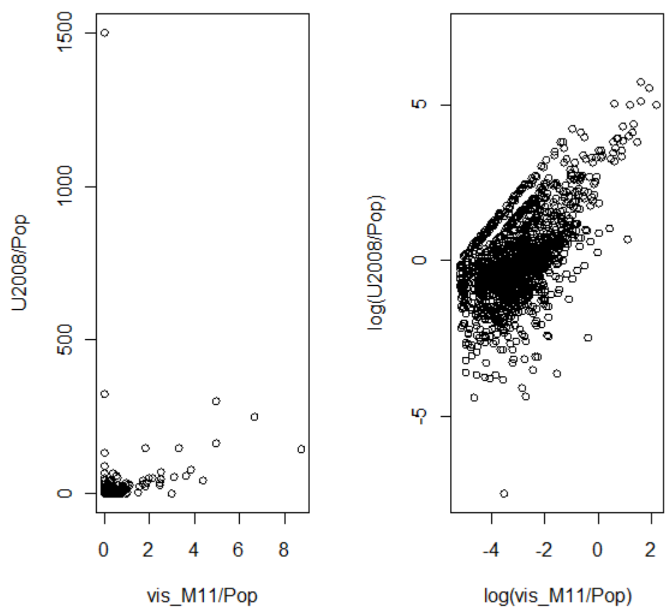



```
#Testing the bivariate srf   wp March09
n=1000
                   #Initialize arrays
sigs = lams =rep(0,n)
acceptmu = 0      #start counters
bols= solve(t(X)%*%X )%*%t(X)%*%y
e=y-X%*%bols;sigy=sd(e)
thx=list(bx=c(0,0),Hxi=diag(2)/1000,s2x=0,nux=1)
lamx=list(sx=3,lam1x=1)
lam = log(mean(y/x) ) #starting value
beta=bols
for(i in 1:n) {
                  #Sampling beta
sb = samplebeta(y,x,X,beta,sig,a=2,thx,lam)
beta =sb$bet
betas= rbind(betas, beta )              #store beta
acceptmu = acceptmu + sb$accept   #counter mu
                                    #Sampling sig
ss = samplesig(y,X,beta,thx)
sig = ss$sig
sigs[i] =   sig             #store sig
ll = samplelam(y,x,beta,sig,lamx)
lam = ll$lam
lams[i] =   lam            #store sig
print(c(i,beta,lam,sig))
  }
apply(betas,2,mean) ; apply(betas,2,sd);mean(acceptmu)
```

7 Sales in pharma marketing

We have observed regional whole sales in Germany for a certain product and we want to relate it to the intensity of doctors visits in a region. The variable a (promotional expenditures) is calculated as number of visits divided by the number of doctors.
The sales response model can be set up in the same way as in the SFR CSSR model. As the first model we look at the sales U2007 and the visits. The scatter plots are seen in figures (**??**) and (**??**) An interesting variance convergence

Fig. 6. Visits M11/M13 and sales U2008 (logs)



relationship is given by the plot (**??**) on per capita Visits M13 and sales U2008 (logs):

The model "visits M11 pc and U2008 pc": The estimates (posterior means and SD) for this model are: $\text{beta}_0(SD) = 1.121(0.0827)$

$\text{beta}_1(SD) = 0.321(0.0222)$

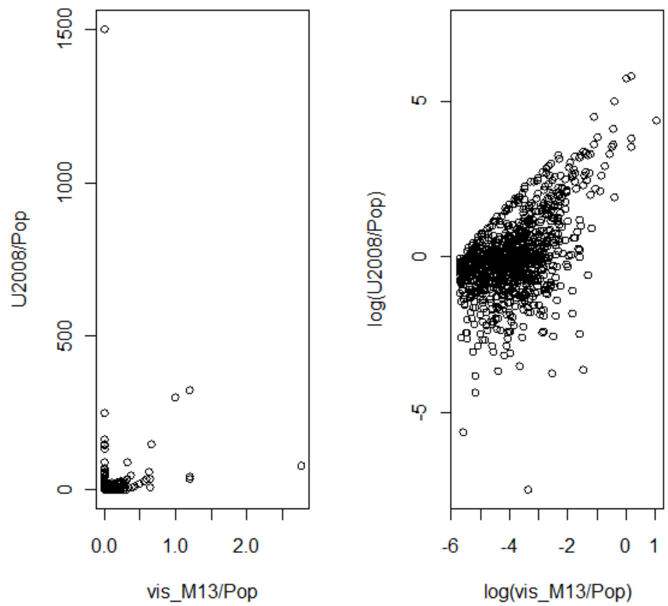
acceptance rate = 50.8

lambda (SD) = 27.8 (1.678)

sigma (SD) = 1.26 (0.021)

```
OLS fit:\\
```

Fig. 7. per capita Visits M13 and sales U2008 pc(logs)




```
Residual Standard Error=1.2581, R-Square=0.1502 \\
F-statistic (df=1, 1898)=335.4101, p-value=0

          Estimate Std.Err t-value Pr(>|t|)
Intercept   1.1237  0.0673 16.6984        0
X           0.3222  0.0176 18.3142        0
```

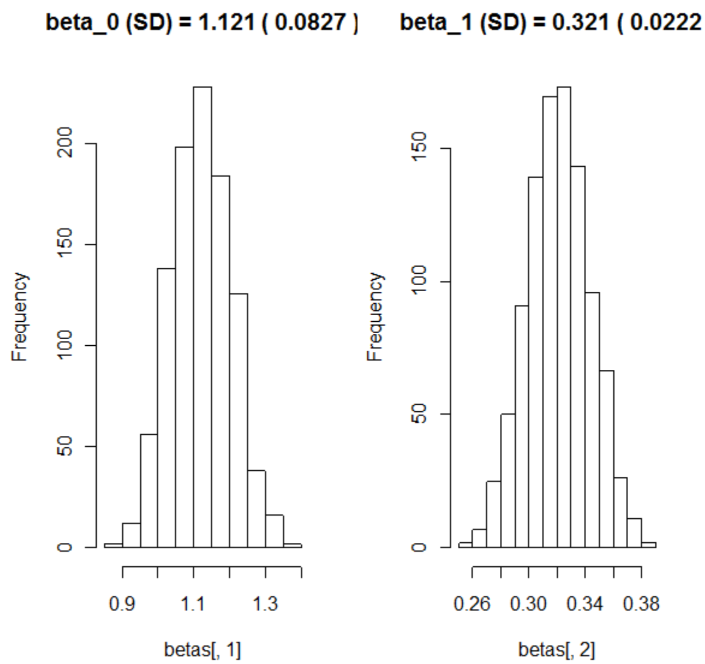
The results of the model "visitsM11 and U2008": The estimates (posterior means and SD) for this model are:

```
beta1: 6.36583618  (0.07469582)
beta2: 0.05815747 (0.02210080)
acceptance rate= 37.6
 mean(SD) of $\lambda$:  8.026314 (2.933747)
 mean(sigs); sd(sigs)    0.9877118 (0.01610788)
```

The model "visits$_M$13$andU$2007" : $TheposteriormeansandSDforthismodelare:$

```
beta_0 (SD) = 6.497 ( 0.0452 )\\
beta_1 (SD) = 0.033 ( 0.0197 )\\
acceptance rate = 50 %
```

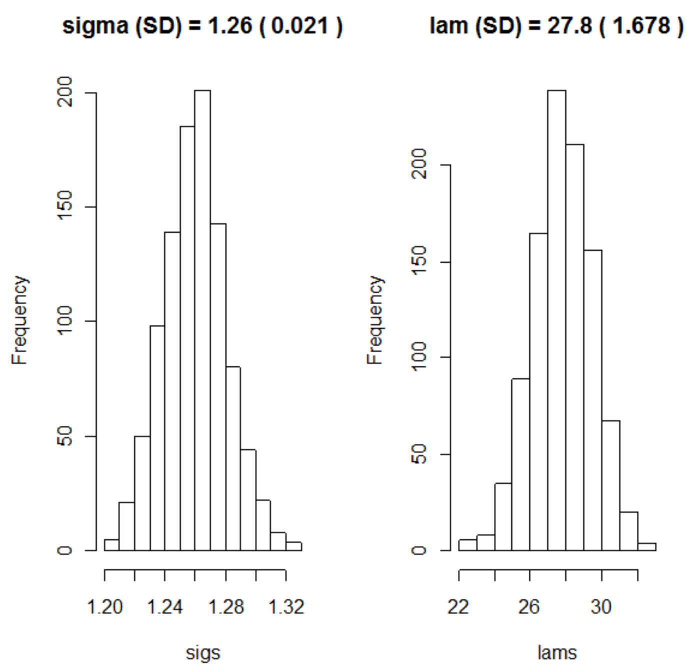
Fig. 8. Betas of visits and sales U2008 per capita (pc)




```
lambda (SD) = 11.753 ( 6.734 )
sigma (SD) =  0.991 ( 0.016 )
```

```
\begin{tabular}{|||||||||||||||}
  \hline
  % after \\: \hline or \cline{col1-col2} \cline{col3-col4} ...
           area vis_M11 vis_M12 vis_M13 visits DocsM  Docs A2007 U2008_M1 U2008_M2   Pop
area       1.00    0.20   -0.40    0.27   0.06  0.08  0.03  0.08    -0.07     0.04 -0.01
vis_M11    0.20    1.00   -0.07    0.29   0.84  0.77  0.09 -0.23    -0.20     0.04  0.16
vis_M12   -0.40   -0.07    1.00   -0.29   0.34  0.33  0.13 -0.23     0.00    -0.05  0.23
vis_M13    0.27    0.29   -0.29    1.00   0.39  0.48  0.17 -0.09    -0.15     0.00  0.13 -
visits     0.06    0.84    0.34    0.39   1.00  0.99  0.25 -0.34    -0.22     0.00  0.33
DocsM      0.08    0.77    0.33    0.48   0.99  1.00  0.29 -0.34    -0.22     0.00  0.36
Docs       0.03    0.09    0.13    0.17   0.25  0.29  1.00 -0.26    -0.10     0.02  0.39 -
A2007      0.08   -0.23   -0.23   -0.09  -0.34 -0.34 -0.26  1.00     0.13     0.04 -0.27
U2008_M1  -0.07   -0.20    0.00   -0.15  -0.22 -0.22 -0.10  0.13     1.00     0.10 -0.12 -
U2008_M2   0.04    0.04   -0.05    0.00   0.00  0.00  0.02  0.04     0.10     1.00  0.00 -
```

Fig. 9. Lam, sig of visits and sales U2008 pc




```
Pop      -0.01    0.16    0.23    0.13    0.33  0.36  0.39 -0.27    -0.12     0.00  1.00 -
PPP       0.00    0.06    0.00   -0.03    0.03  0.02 -0.28  0.02    -0.02    -0.02 -0.32
U2008     0.03    0.12    0.04    0.00    0.11  0.10  0.07  0.36     0.06     0.07  0.01 -
U2007     0.02    0.52    0.33    0.18    0.66  0.65  0.49 -0.38    -0.17     0.03  0.37 -

  \hline
\end{tabular}
```

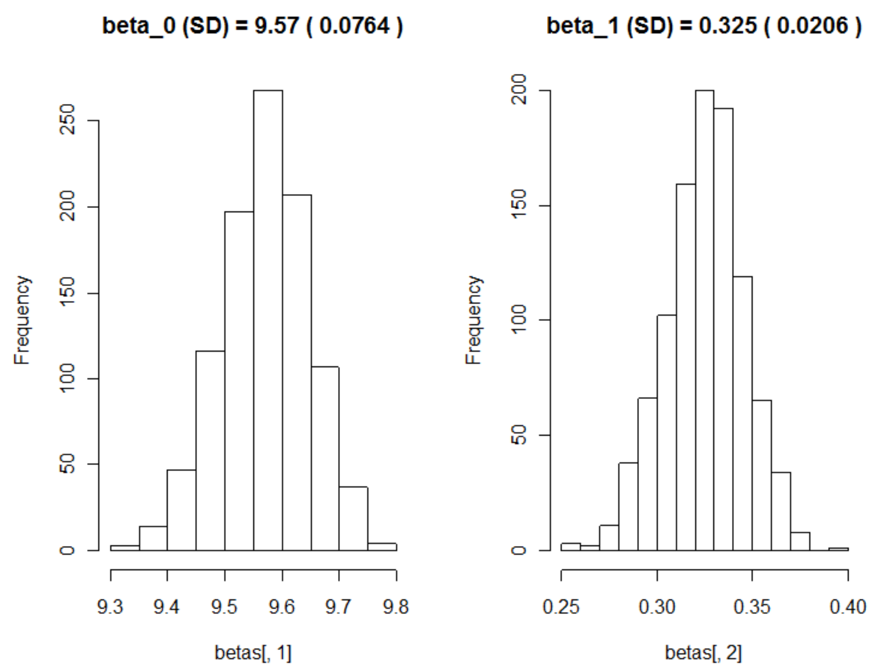
8 Extensions with exogenous variables

In this section we estimate the model as before but we include 2 more regressors: "$visits_M 13$ " and PPP. The correlation with the PPP regressor is quite non-linear and negative, see figure (14):

The Bayesian SRF estimation results are:

```
beta_0 (SD) = 1.152 ( 0.2507 )
beta_1 (SD) = 0.22 ( 0.0251 )
beta_2 (SD) = 0.19 ( 0.0239 )
beta_3 (SD) = 0.19 ( 0.0359 )
```

Fig. 10. visits M11 and sales U2008

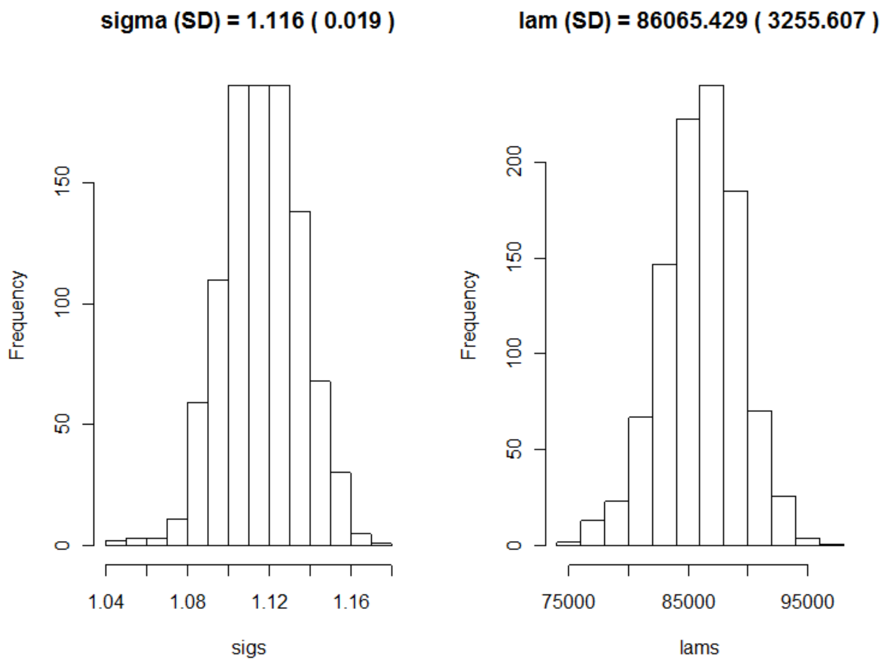



```
acceptance rate = 51 %

lambda (SD) = 22.478 ( 2.268 )
sigma (SD) =  1.225 ( 0.02 )

The OLS fit is [ ls.print(lsfit(cbind(l11pc,l13pc,lppp),y)) ]

Residual Standard Error=1.2238, R-Square=0.1967
F-statistic (df=3, 1896)=154.7628, p-value=0

          Estimate Std.Err t-value Pr(>|t|)
Intercept   1.1510  0.1987  5.7925    0e+00
l11pc       0.2197  0.0197 11.1394    0e+00
l13pc       0.1893  0.0191  9.9163    0e+00
lppp        0.0962  0.0283  3.3938    7e-04
```

The distribution of the parameters are shown in the next figures:

The correlation matrix for the log-transformed variables is

Fig. 11. visitsM11 and sales U2008



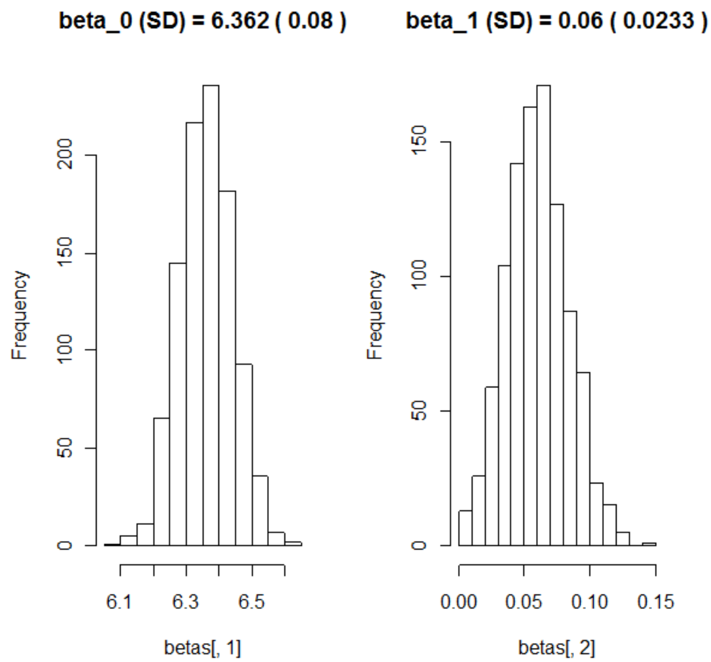
	l2008pc	l2007pc	l11pc	l12pc	l13pc	ldocs	ldocsm	a2007
l2008pc	1.00	0.74	0.39	0.35	0.37	0.68	0.62	0.16
l2007pc	0.74	1.00	0.36	0.37	0.36	0.69	0.61	0.09
l11pc	0.39	0.36	1.00	0.18	0.49	0.47	0.74	0.09
l12pc	0.35	0.37	0.18	1.00	-0.02	0.49	0.50	0.05
l13pc	0.37	0.36	0.49	-0.02	1.00	0.51	0.62	-0.04
ldocs	0.68	0.69	0.47	0.49	0.51	1.00	0.82	-0.04
ldocsm	0.62	0.61	0.74	0.50	0.62	0.82	1.00	0.01
a2007	0.16	0.09	0.09	0.05	-0.04	-0.04	0.01	1.00

9 2 endogenous variables in the SRF

In this section we consider the SRF with 2 endogenous variables:

$$y = \gamma x_1^{\beta_1} x_2^{\beta_2} e^{\epsilon} \tag{48}$$

or taking logs we find for $\beta = (\beta_0, \beta_1, \beta_2)'$ with $\beta_0 = log(\gamma)$



$$ln\ y \sim N[\mu_y = X\beta, \sigma_y^2 I_n] \tag{49}$$

If both input regressors x_1 and x_2 are measure on the same scale we can make the following SPD assumption:

$$\begin{aligned} \partial y/\partial x_1 &= dy_1 = \beta_0\beta_1 x^{\beta_1-1} z^{\beta_2} \\ \partial y/\partial x_2 &= dy_2 = \beta_0\beta_2 x^{\beta_1} z^{\beta_2-1} \end{aligned} \tag{50}$$

and we impose stochastic partial derivatives (SPD) constraints in the following form:

$$\begin{aligned} log(dy_1) &= log\frac{y\beta_1}{x_1} \sim N[\lambda_1, \tau_1^2] \\ log(dy_2) &= log\frac{y\beta_2}{x_2} \sim N[\lambda_2, \tau_2^2] \end{aligned} \tag{51}$$

This defines a simultaneous equation system in the endogenous variables. A special case is obtained if we set with $\lambda_1 = \lambda_2$.

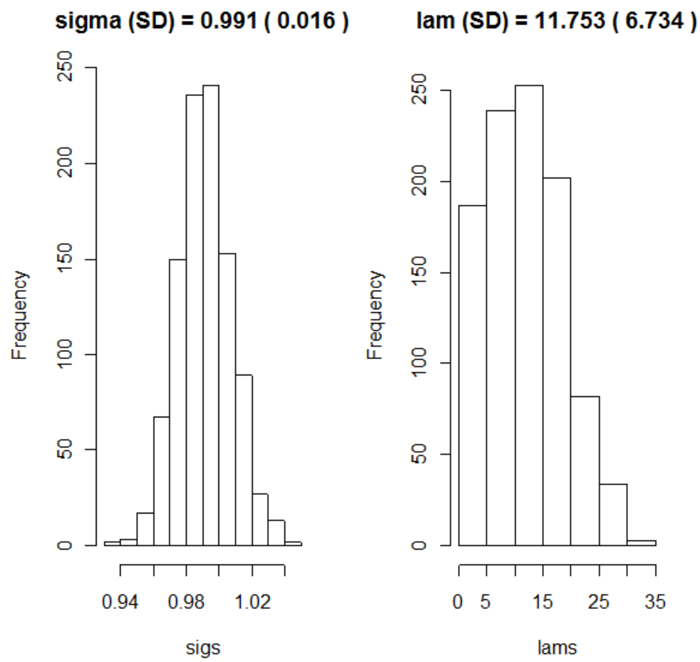


Fig. 14. Betas of PPP and sales U2008pc

Let us assume for the moment that the β_i coefficients and the regressors x_i are fixed. Then we can determine the variances τ_i^2 in (51) by

$$log(y_i) - log(x_i) + log(\beta_i) = \eta_i \quad \text{with} \quad \eta_i \sim N[\lambda_i, \tau_i^2], \quad i = 1, 2; \qquad (52)$$

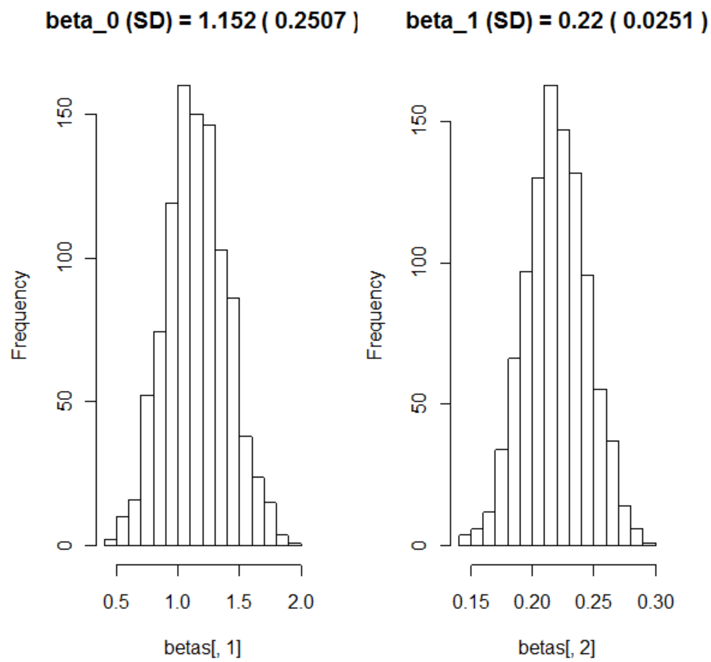
with the moments $E\eta_i = \lambda_i$ and

$$Var(\eta_i) = \tau_i^2 = \sigma_i^2, \quad i = 1, 2 \qquad (53)$$

The 2 partial derivatives constitute now a simultaneous regression system - for fixed β coefficients - in the x and λ variables. By writing the 2 derivatives in matrix form we can make the following stochastic normal assumption about the sizes of the individual derivatives:

$$\lambda = \gamma_0 + Bx \sim N[\mu_\lambda, \Sigma_\lambda] \qquad (54)$$

where μ_λ and Σ_λ are the unknown parameters of the observable stochastic partial derivatives (SPD). Now we define the vectors



$$\gamma_0 = \begin{pmatrix} ln\ \beta_0 + ln\ \beta_1 \\ ln\ \beta_0 + ln\ \beta_2 \end{pmatrix} \text{ and } \quad \gamma = B^{-1}\gamma_0. \tag{55}$$

which are needed if we express the endogenous regressor variables as functions of the SPD's:

$$x = B^{-1}(\lambda - \gamma_0) \text{or} \begin{pmatrix} ln\ x_1 \\ ln\ x_2 \end{pmatrix} = \begin{pmatrix} \beta_1 - 1 & \beta_1 \\ \beta_2 & \beta_2 - 1 \end{pmatrix}^{-1} \begin{pmatrix} ln\ \lambda_1 - ln\ \beta_0 - ln\ \beta_1 \\ ln\ \lambda_2 - ln\ \beta_0 - ln\ \beta_2 \end{pmatrix}$$

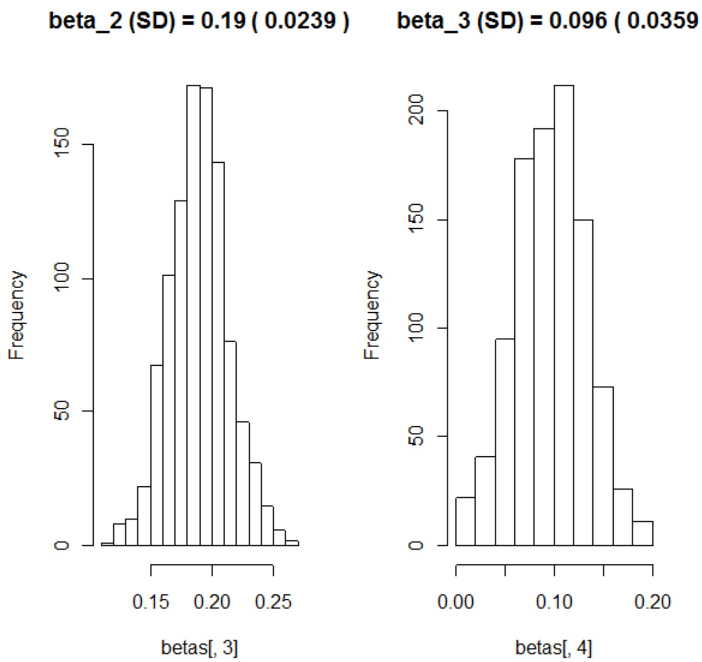
The Jacobian J is just the determinant of the B^{-1} matrix: $J = |B|^{-1} = [(\beta_1 - 1)(\beta_2 - 1) - \beta_1\beta_2]^{-1}$ with

$$B = \begin{pmatrix} \beta_1 - 1 & \beta_1 \\ \beta_2 & \beta_2 - 1 \end{pmatrix} \tag{56}$$

Next we have to find the variance of the endogenous $ln\ x = (ln\ x_1, ln\ x_2)'$ variable vector

$$Var(ln\ x \mid \beta) = Var(B^{-1}\lambda) = B^{-1}\Sigma_\lambda B^{-1\prime} = (B'\Sigma_\lambda^{-1}B)^{-1} \tag{57}$$

and the therefore the whole distribution is normal:



$$ln\ x \mid \beta \sim N[\mu_x = \gamma + B^{-1}\lambda, \Sigma_x = (B'\Sigma_\lambda^{-1}B)^{-1}] \tag{58}$$

The fcd for the $\lambda = (\lambda_1, \lambda_2)'$ parameters are proportional to

$$p(\lambda \mid \theta^c) \propto N[\lambda_*, \Sigma_{\lambda,*}]l(\lambda \mid y) \propto N[\lambda_{**}, \Sigma_{\lambda,**}] \tag{59}$$

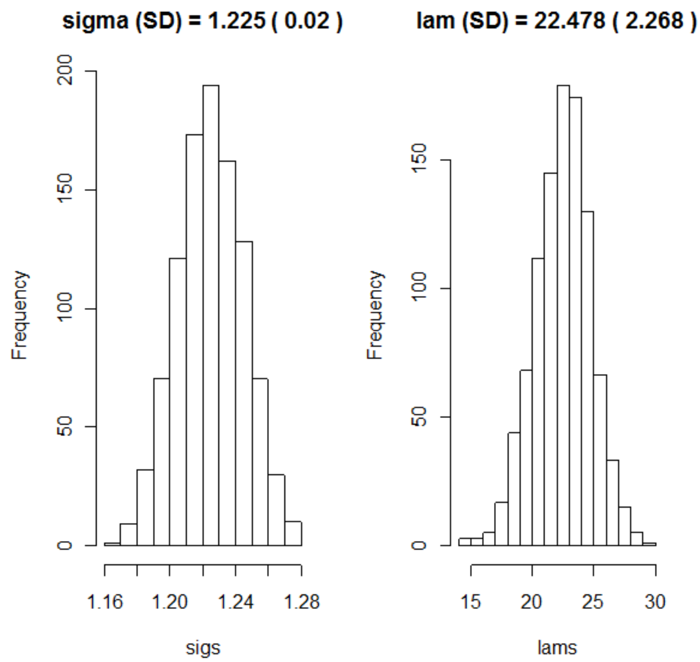
where θ^c are the remaining parameters and the likelihood function $l(\lambda \mid y)$ is given by (51,52) in the following way by the moments of the stochastic derivatives:

$$E(dy_{(i)} \mid \theta^c) = \lambda_i \tag{60}$$

$$\begin{aligned} \sigma^2_{dy,i} &= Var(dy_{(i)} \mid \theta^c) = Var(log(y_i) - log(x_i)) = \qquad (61)\\ &= Var(log(y_i) + Var(log(x_i) - 2Cov(log(x_i), log(y_i)) = \\ &= \sigma_y^2 + \sigma_{x,ii} - 2\sigma_{xy,i}, \\ &= \sigma_y^2 + \sigma_{x,ii} - 2\beta_i\sigma_{x,ii} \\ &= \sigma_y^2 + \sigma_{x,ii}(1 - 2\beta_i) \quad i = 1, 2 \end{aligned}$$

because of (49), (58) with

Fig. 17. Lam, sigma of the 2+2 model on sales U2008pc



$$\Sigma_x = \begin{pmatrix} \sigma_{x,11} & \sigma_{x,12} \\ \sigma_{x,21} & \sigma_{x,22} \end{pmatrix} \tag{62}$$

and the covariance term is

$$\begin{aligned} Cov(log(x_i), log(y_i)) &= \sigma_{xy,i} = E[(log(x_i) - \mu_i)(log(y_i) - \mu_y)] \qquad (63) \\ &= E[(log(x_i) - \mu_i)(... + \beta_1 log(x_1) + \beta_2 log(x_2) + ... - \mu_y)] \\ &= \beta_i \sigma_{x,ii} \end{aligned}$$

with μ_y given in (49) and μ_i in (58). Finally we can write briefly $dy_{(i)} \mid \theta^c \sim N[\lambda_i, \sigma^2_{dy,i}]$ for $i = 1, 2$.
If we assume independence between the 2 stochastic derivatives we can estimate each λ_i separately by a univariate Bayesian normal model. For the n observed values of the i-th stochastic derivatives we use the formula (52) and the fcd's for the λ_i are given by a conjugate normal density:

$$\tau^{-2}_{i**} = \tau^{-2}_{i*} + \sigma^{-2}_{dyi} n \tag{64}$$

for $i = 1, 2$ and

$$\lambda_{i**} = \tau^2_{i**}[\tau^{-2}_{i*}\lambda_{1*} + \sigma^{-2}_{dyi} n \mu_{\lambda_i}], \tag{65}$$

with $\mu_{\lambda_i} = \frac{1}{n}\sum_j dyi_j$ being the observed mean of the i-th stochastic derivatives and $\sigma^2_{dyi} = \frac{1}{n}\sum_j (dyi_j - \mu_{\lambda_i})^2$ the observed variance of the i-th SPD.
Note that a joint bivariate estimation of the λ_i's is possible and also a pooled estimation if we assume only one common $\lambda_i = \lambda$ parameter.

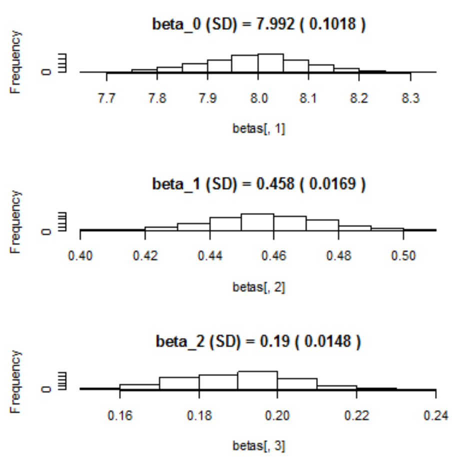
$$\tau_{**}^{-2} = \tau_*^{-2} + \sigma_{dy1}^{-2} n + \sigma_{dy2}^{-2} n \tag{66}$$

$$\lambda_{**} = \tau_{**}^2 [\tau_*^{-2}\lambda_* + \sigma_{dy1}^{-2} n\mu_{\lambda_1} + \sigma_{dy2}^{-2} n\mu_{\lambda_2}]. \tag{67}$$

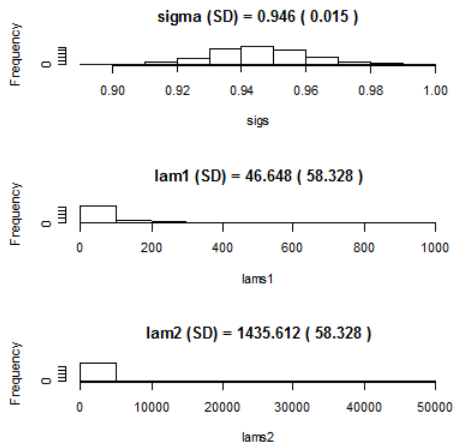
9.1 Example with 2 edogenous variables

We consider the pharma sales examples again, but now we estimate the whole-sales (y) by two visits of 2 different products (x11 and x13) from the same company. Is there a joint optimal effect visible in the estimates?
The coefficient estimates are

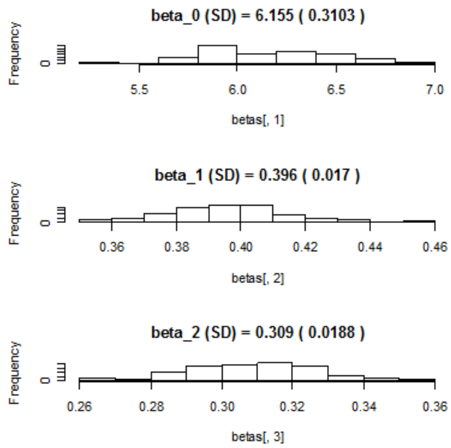
```
 beta_0 (SD) = 7.992 ( 0.1018 )
beta_1 (SD) = 0.458 ( 0.0169 )
beta_2 (SD) = 0.19 ( 0.0148 )
acceptance rate = 49 %
sigma (SD) =  0.946 ( 0.015 )
lambda1 (SD) = 46.648 ( 58.328 )
lambda2 (SD) = 1435.612 ( 2577.229 )
```

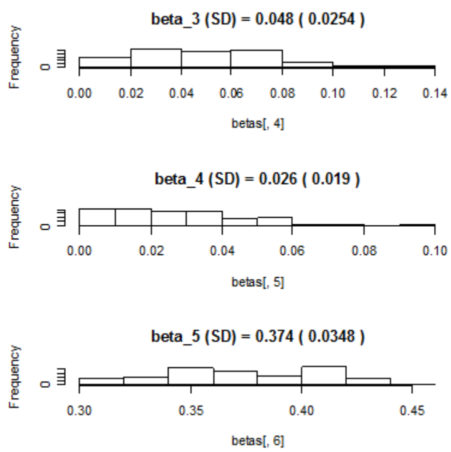


With 3 more exogenous variables log U2007 (previous year sales), logPPP (purchasing power potential) and logPop (population) we get the following result




```
beta_0 (SD) = 6.155 ( 0.3103 )
beta_1 (SD) = 0.396 ( 0.017 )
beta_2 (SD) = 0.309 ( 0.0188 )
acceptance rate = 53 %
sigma (SD) =  0.969 ( 0.097 )
lambda1 (SD) = 40.568 ( 51.176 )
lambda2 (SD) = 1675.359 ( 1664.061 )
OLS:
Residual Standard Error=0.8928,R-Square=0.9891
F-statistic (df=6, 1894)=28635.66 ,p-value=0
      Estimate Std.Err t-value Pr(>|t|)
        6.2078  0.2463 25.2089   0.0000
x11     0.3956  0.0152 26.1093   0.0000
x13     0.3100  0.0148 20.9276   0.0000
u2007   0.0463  0.0207  2.2381   0.0253
ppp    -0.0248  0.0209 -1.1881   0.2349
pop     0.3713  0.0253 14.6627   0.0000
```

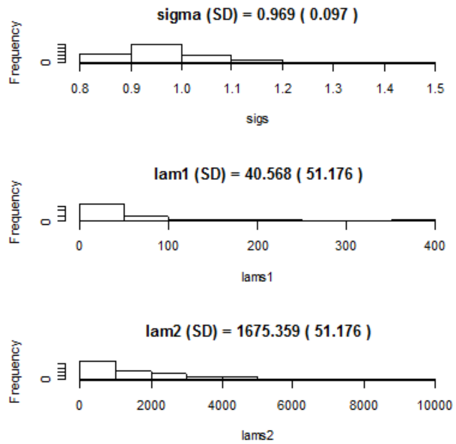




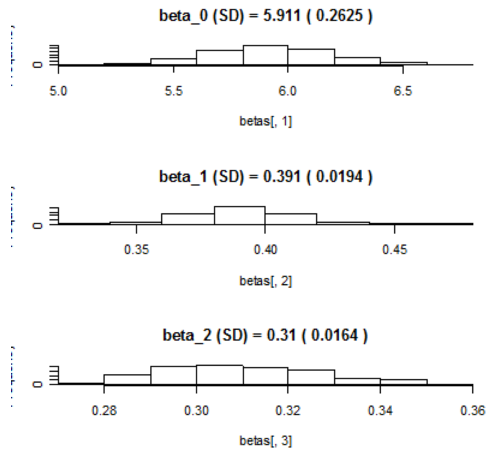
10 with growth Dummy variable

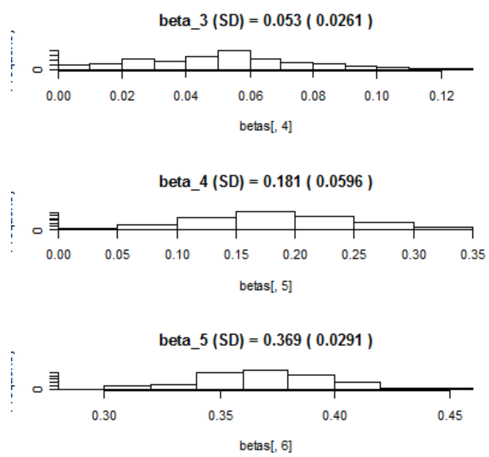
We define a dummy variable Dposg which splits the observation into 2 groups: positive or negative total sales growth in the year 2007/2008. (if(g78[i] ¿0) Dposg[i] =1)

1. MARKET SHARE



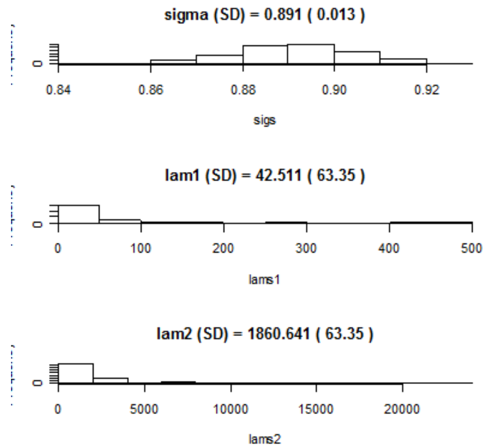

```
beta_0 (SD) = 10.827 ( 0.5485 )
beta_1 (SD) = 0.701 ( 0.0443 )
beta_2 (SD) = 0.421 ( 0.0493 )
acceptance rate = 51 %
sigma (SD) =  6.928 ( 0.834 )
lambda1 (SD) = 211.75 ( 406.737 )
lambda2 (SD) = 1592.964 ( 3476.425 )
OLS
Residual Standard Error=1.5038. R-Square=0.9829
F-statistic (df=6, 1894)=18180.42, p-value=0

      Estimate Std.Err  t-value Pr(>|t|)
       10.8503  0.3351  32.3750        0
x11     0.6986  0.0255  27.3926        0
x13     0.4203  0.0250  16.8147        0
u2007  -0.4202  0.0350 -12.0169        0
Dposg  -0.7949  0.0805  -9.8682        0
pop     0.5762  0.0421  13.6733        0

beta_3 (SD) = 0.423 ( 0.0605 )
beta_4 (SD) = 0.761 ( 0.1423 )
beta_5 (SD) = 0.578 ( 0.0841 )
```



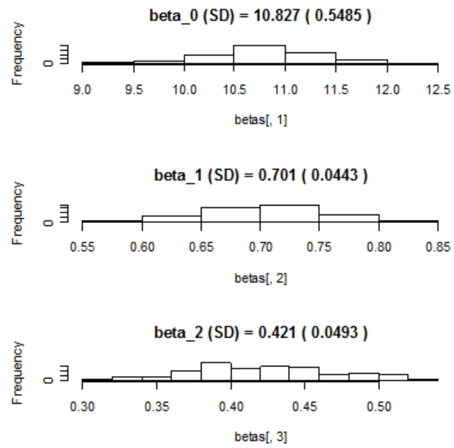


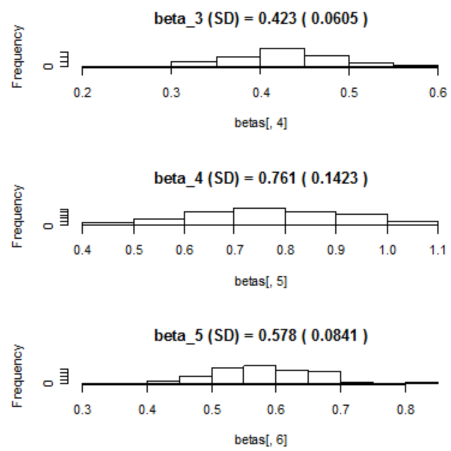
2. M1 sales

```
beta_0 (SD) = 5.911 ( 0.2625 )
beta_1 (SD) = 0.391 ( 0.0194 )
beta_2 (SD) = 0.31 ( 0.0164 )
acceptance rate = 49.5 %
sigma (SD) =  0.891 ( 0.013 )
lambda1 (SD) = 42.511 ( 63.35 )
```



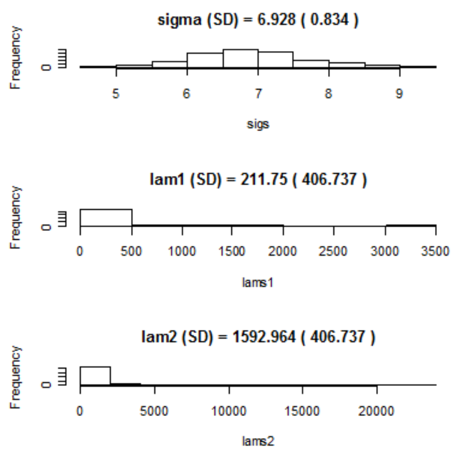

```
lambda2 (SD) = 1860.641 ( 2827.231 )
beta_3 (SD) = 0.053 ( 0.0261 )
beta_4 (SD) = 0.181 ( 0.0596 )
beta_5 (SD) = 0.369 ( 0.0291 )
OLS
Residual Standard Error=0.8897, R-Square=0.9892
F-statistic (df=6, 1894)=28838.04, p-value=0

      Estimate Std.Err t-value Pr(>|t|)
        5.9080  0.1983 29.7938   0.0000
x11     0.3910  0.0151 25.9110   0.0000
x13     0.3064  0.0148 20.7195   0.0000
u2007   0.0529  0.0207  2.5559   0.0107
Dposg   0.1825  0.0477  3.8290   0.0001
pop     0.3665  0.0249 14.6987   0.0000
```





11 Summary

In this paper we have developed a spatial extension for a cross-sectional sales response model which obeys stochastic constraints, as it was suggested by Kao et al.(2005). Again, a simple MCMC estimation of the model turns out to be



quite straightforward, despite the many constraints. Nevertheless, the SRF with stochastic partial derivative (SPD) constraints leave many problem of the appropriate estimation method and for model choice open. In a simulation study we have shown that the estimation without the optimality constraints create biases in the coefficient estimates. The SRF can be extended to a spatial cross-sectional sales response model that takes the neighborhood structure of the observations into account. We demonstrate this approach by a regional pharmaceutical sale models for Germany.

12 Appendix: MCMC in the SAR model

[MCMC in the Normal-Gamma SAR model]

1. Starting values: set $\beta = \beta_{OLS}$ and $\rho = 0$
2. Draw σ^{-2} from $\Gamma[\sigma^{-2} \mid s^2_{**}, n_{**}]$
3. Draw β from $\mathcal{N}[\beta \mid \mathbf{b}_{**}, \mathbf{H}_{**}]$
4. Draw ρ from $p(\rho \mid \beta, \sigma^{-2})$
5. Repeat until convergence.

The full conditional distributions are:

1. For the regression coefficients β

$$p(\beta \mid \sigma^{-2}, \rho, \mathbf{y}) \propto \mathcal{N}[\beta \mid \mathbf{b}_{**}, \mathbf{H}_{**}] \tag{68}$$

with the hyper-parameters

$$\mathbf{H}_{**}^{-1} = \mathbf{H}_{*}^{-1} + \sigma^{-2}\mathbf{X}\mathbf{X}, \tag{69}$$

$$\mathbf{b}_{**} = \mathbf{H}_{**}\left[\mathbf{H}_{*}^{-1}\mathbf{b}_{*} + \sigma^{-2}\mathbf{X}\mathbf{R}\mathbf{y}\right]. \tag{70}$$

2. We find for the residual variance σ^2

$$p(\sigma^{-2} \mid \beta, \rho, \mathbf{y}) \propto \Gamma[\sigma^{-2} \mid s^2_{**}, n_{**}], \tag{71}$$

a gamma distribution with the parameters

$$n_{**} = n_* + n, \tag{72}$$
$$n_{**}s^2_{**} = n_* s^2_{**} + (\mathbf{Ry} - \mathbf{X}\beta)(\mathbf{Ry} - \mathbf{X}\beta). \tag{73}$$

3. The full conditional distribution for the spatial correlation coefficients ρ is proportional to

$$p(\rho \mid \beta, \sigma^{-2}, \mathbf{y}) \propto |\mathbf{I}_n - \rho\mathbf{W} \mid \exp\left(-\frac{\epsilon'_\rho \epsilon_\rho}{2\sigma^2}\right) \tag{74}$$

where the residuals of the spatial regression are

$$\epsilon_\rho = \mathbf{y} - \mathbf{X}\beta - \rho\mathbf{W}\mathbf{y}$$

. We sample ρ_t^{new} from a normal proposal:

$$\rho_t^{new} = \rho_t^{old} + c\phi, \quad \phi \sim \mathcal{N}[0, 1]. \tag{75}$$

The scalar c is a tuning parameter and ρ^{old} is the parameter of the previous sampling step. Next, we evaluate the acceptance probability

$$\alpha(\rho_t^{old}, \rho_t^{new}) = min\left(\frac{p(\rho_t^{new})}{p(\rho_t^{old})}, 1\right), \tag{76}$$

where p is the full conditional distribution in (**??**) and, of course, ρ_t in ϵ_ρ also changes to ρ_t^{new} in $p(\rho_t^{new})$. Finally we set $\rho_t = \rho_t^{new}$ with probability $\alpha(\rho_t^{old}, \rho_t^{new})$, otherwise $\rho_t = \rho_t^{old}$.

The scalar c is tuned to produce an acceptance rate between 10% and 30% as is suggested in Holloway *et al.* (2002). It should be mentioned that the proposal density of ρ_t is not truncated to the interval $(\lambda^{-1}_{min}, \lambda^{-1}_{max})$ since the constraint is part of the target density. Thus, if the proposal value of ρ is not within the interval, the conditional posterior is zero, and the proposal value is rejected with probability one (see Chib and Greenberg, 1998).
The likelihood function is given by

$$p_{\mathcal{N}}(\mathbf{y} \mid \beta, \sigma^{-2}, \rho, \mathbf{X}, \mathbf{W}) = (2\pi\sigma^2)^{-1/2}|\mathbf{I}_n - \rho\mathbf{W}| \tag{77}$$
$$\cdot \ \exp\left(-\frac{(\tilde{\mathbf{y}} - \mathbf{X}\beta)(\tilde{\mathbf{y}} - \mathbf{X}\beta)}{2\sigma^2}\right) \tag{78}$$

The marginal likelihood is given by the Newton-Raftery formula

$$p_{\mathcal{N}}(\mathbf{y} \mid ...)^{-1} = \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} p_{\mathcal{N}}(\mathbf{y} \mid \theta_{(i)}, \mathbf{X}, \mathbf{W})^{-1} \tag{79}$$

with the parameters given for simulation i by $\theta_{(i)} = (\beta_{(i)}, \sigma^{-2}_{(i)}, \rho_{(i)})$ and the likelihood $p_{\mathcal{N}}(\mathbf{y} \mid \theta_{(i)}, \mathbf{X}, \mathbf{W})$.

References

1. L.-J. Kao, C.-C. Chiu, T.J. Gilbride, T. Otter, and G.M. Allenby. Evaluating the Effectiveness of Marketing Expenditures. *Working Paper*, Ohio State University, Fisher College of Business, 2005.
2. P.E. Rossi, G.M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing.* John Wiley and Sons, New York, 2005.