



WP 11-50

Wolfgang Polasek

Institute for Advanced Studies, Austria

University of Porto, Portugal

The Rimini Centre for Economic Analysis (RCEA), Italy

MARKETING RESPONSE MODELS FOR SHRINKING BEER SALES IN GERMANY

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, nonprofit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *The Rimini Conference in Economics and Finance* (RCEF). The RCEA has a Canadian branch: *The Rimini Centre for Economic Analysis in Canada* (RCEA-Canada). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers and Professional Report series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

The Rimini Centre for Economic Analysis

Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy

www.rcfea.org - secretary@rcfea.org

Marketing response models for shrinking beer sales in Germany

Wolfgang Polasek¹

IHS, Stumpergasse 56, A-1060 Vienna, and CMUP, Porto polasek@ihs.ac.at

Summary. Beer sales in Germany are confronted for several years with a shrinking market share in the market of alcoholic beverages. I use the approach of sales response function (SRF) models as in Polasek and Baier (2010) and adapt it to time series observation of beer sales for simultaneous estimation. I propose a new class of growth sales (gSRF) models having endogenous and exogenous variables as in Polasek (2011) together with marketing efforts that follow a sustained growth allocation principle. This approach allows to model growth rates in markets that are exposed to fierce competition and where marketing efforts cannot be evaluated directly. The class of gSRF models has the property that it models supply (i.e. marketing efforts) and demand factors jointly in a log-linear regression model that are correlated over time. The estimated model can explain the relative success of marketing expenditures for the shrinking beer market in the period 1999-2010.

Key words: Sales response functions (SRF), marketing budget models, MCMC estimation, beer consumption, optimal budget allocation.

1 Introduction

Beer consumption and production in Germany is an important economic activity but has declined over the last decade. Therefore it is rather surprising that regional beer consumption is not available as a panel data set from the German statistical office. Only marginal data, like the total beer production or the total marketing effort per year is available. This incomplete data base was the starting point of this paper: Is it possible to make inference in a short time series model even if detailed information across regions is missing and the marketing strategies of the many (regional) beer companies are not known?

Kao et al. (2005) have proposed a simultaneous estimation of marketing success in dependence of optimal inputs in a sales response function (SRF) model. The main idea behind this approach is that the (optimal) expenditures for inputs might depend on the current sales and should be estimated endogenously.

Polasek (2010) has introduced a family of multiplicative SRF(k) for a cross-sectional sample where the parameter k denotes the number of input variables (sales expenditures and sales related covariates) that are producing the sales output by a Cobb-Douglas type of production function. The multiplicative model is extended to an additive SRF(k) model and a semi-additive SRF(k) model, where we have a mixture of additive and multiplicative input terms.

The current approach emphasizes the system approach of the demand-supply system for the estimation of a SRF in a panel model, because the input variables (marketing efforts) are jointly determined by the output (sales). This approach is the focus of macroeconomics and developed in econometrics since several decades. New is the assumption that the endogeneity of the inputs stem from an implied (stochastic) optimality consideration, which is imposed through a first derivative constraint.

The paper is laid out as follows. In the next section 2 we justify our approach to sales response models by some general considerations. In section 3 we describe the basic SRF(1)-SPD model and the estimation approach. Section 4 extends this approach to SRF(1)-AR models, since we have to expect in time series models auto-correlated errors. In section 5 we discuss the Example for the German beer market and the final section concludes.

1.1 Other approaches to beer marketing

There are almost no studies that try to quantify the impact of advertisement on beer sales, also not on an international comparison. This might change if the beer industry becomes a more global player and has to market beer brands internationally. If there are more mergers and acquisitions, then this will change the marketing strategies. In an article of 'Marketwatch' (<http://www.marketwatch.com/story/inbev-takeover-spotlights-anheuser-buschs-big-ad-budget>) we find the following quote in reaction to the takeover of Anheuser-Busch: "Will sports lose one of its biggest boosters? InBev takeover spotlights Anheuser-Busch's big ad budget. . . . That leaves Anheuser-Busch's massive advertising and sponsorship budget perhaps the juiciest target of swift cost-cuts.

The No.1 U.S. beer maker is the nation's 22nd largest advertiser, according to data compiled by Advertising Age and TNS Media Intelligence, with total expenditures of \$1.36 billion last year. About a third of that – \$475 million – is spent on TV, radio, magazines and the Internet, with the rest aimed at trade promotions, sponsorships, point-of purchase ad space and the like.

By contrast, SAB-Miller spent \$230 million on U.S. media last year and Diageo, the world's largest spirits company with revenues in excess of Anheuser-Busch's, laid out \$173 million. . . ."

Thus, in the aftermath of market concentration, new strategies on spending ad budgets will be developed. This will also affect the sponsoring market, as the following quote on the same web site shows:

”Will sports lose its biggest ad booster? ... How will InBev’s \$52 billion takeover of Anheuser-Busch affect the No. 1 U.S. brewer’s massive sports advertising and sponsorship budget?”

2 Some remarks on SRF models and their estimation

A shrinking sales market like the German beer market is challenge for quantitative marketing models. The recent class of sales response models that were promoted by Kao et al. (2005) and Baier and Polasek (2010) provide a flexible framework to estimate by MCMC models the reaction of marketing efforts in such markets. Models for sales response functions (SRF) provide a class of models that can be combined with many additional assumptions on the relationship between the supply and demand side of the market. For example, in cross-sections or panels, the SRF model can be extended to the SRF-SAR model, where SAR stands for a spatial autoregressive model (see e.g. Anselin 1986), as will be shown in a next paper.

Common to all approaches is that they assume a behavioral model for sales and marketing actions and therefore an appropriate joint model for the demand and the supply side has to be found. The supply side model is governed by an market strategy that follows some optimality principle. For the shrinking beer market we have assume a new sustained optimal growth allocation principle that results in an allocation rule that sales expenditures follow a constant rule that is proportional to the first partial derivative over time. Since this variable cannot be observed directly we need to assume a latent variable.

The estimation of the latent variable SRF model for the German beer market shows that there has been some positive effects of beer marketing expenditures to fight the shrinking markets that was mainly caused by the increasing market share of the wine sales in the German alcoholic beverage market. Therefore we need to estimate an SRF(1)X(1) model where the X stand for the exogenous (control) variable, in our case the increasing wine share in the market.

For this shrinking sales over time we suggest to use a multiplicative SRF model applied to growth factors, briefly denoted by gSRFX model. Also we cannot use the Albers marketing allocation rule as a response of the supply side, since no regional data are involved. Instead we suggest that the supply side follows a sustained growth allocation rule of the marketing expenditures, such that a simple allocation rule follows. In such behavioral models it is not necessary to assume that marketing strategist will do in practice these mathematical calculations, rather we like to know if such a combination of simple SRF models and observed marketing expenditures lead to a simultaneous model of right and left hand side variables that explains the reality better. Because of the complexity of the parameters involved and the usually small

data base it seems that MCMC methods work best at the moment to achieve this goal.

2.1 Stochastic partial derivatives (SPD) in a sales model

If the first derivative of a sales response (SRF) model is used as a latent variable then the sales equation (log-y equation) and the derivative equation imply a simultaneous equation system, and the stochastic allocation restriction imply an endogeneity of the single input variable x . The following 3 stochastic assumptions are the basic building blocks for the SRF-SPD model and are based on 3 types of considerations that reflect the interaction of the actions observed in the market from the demand and the supply side that leads to appropriate steps (= equations) in the joint model building process:

1. The stochastic (demand) model: input variables and the functional form imply (\Rightarrow) output variables plus noise.
2. The allocation model (supply side): Stochastic response model + imposing optimal expenditure allocations = input variables & functional form (SRF) \Rightarrow first derivatives plus noise.
3. The final model (conditional on assumed demand and supply responses): Known SRF coefficients & SPD assumptions \Rightarrow stochastic regressors (pivotal variable change).

The sales response model that assumes stochastic SPD allocations implies the following additional (implicit) assumptions that are part of the estimation process:

1. The "stochastic ads allocation" rule: We assume that the realized derivatives of sales w.r.t. marketing efforts are approximately equal. We use the concept of realized derivatives to emphasize the fact, that the exact sales changes are unknown and have to be estimated for the estimation of the model by the first derivatives, which is model dependent, i.e. depends on the assumed functional form of the SRF. The company management have learned in the past to understand and to know about the SRF function in their field, even if the exact functional form is unknown to them. Therefore they use the input variables in an optimal way, that is they look to spend promotional money in such a way that for each region the change in sales is about equal.
2. This implied behavioral assumption of the SRF model has to be incorporated into the estimation process and leads to a larger model class of system estimation, since the input and output variables (y and x) are endogenously linked by this assumption.
3. Thus the SDP assumption (for ads allocations) implies a joint distribution of all the endogenous variables, since the realized derivatives depend on the functional form of the SRF model.
4. The derivative w.r.t. marketing expenditures cannot be directly observed

(either the amount channeled through the input variables is unknown or the sales changes are not reported on this disaggregated level or is imprecisely measured). Thus, it becomes necessary to introduce the unobserved derivative as a latent variable in the estimation process.

5. In the MCMC estimation procedure the latent derivatives are generated by so-called 'direct simulations' from the current specification of the SRF model.

6. The latent variable can be viewed as a proxy variable, which is simulated through a model that uses the exogenous regressors of the system.

In the next section we introduce the SRF(1) model and the MCMC estimation under the SPD assumption.

In Section 5 we discuss a regional sales response model that involves data from the German beer market for the period 1997 to 2010. In a final section we conclude.

3 The SRF(1) model with optimal allocations (OA)

In this section we start with the simple SRF(1) model because we want to demonstrate the consequences of the OA and SPD assumption for the estimation procedure.

We consider the SRF(1) sales response function $y = y(x)$ with one input variable x

$$y = \beta_0 x^{\beta_1} e^\epsilon, \quad (1)$$

where ϵ is assumed to be a $N[0, \sigma_y^2]$ distributed error term. By taking logs for the n cross-sectional observations we find the following linear regression model

$$\ln y \sim N[\mu_y = X\beta, \sigma_y^2 I_n] \quad (2)$$

with the regression coefficients $\beta = (\ln \beta_0, \beta_1)$ and the regressor matrix $X = (1_n : \ln x)$ where 1_n is a vector of 1's and x is the cross-sectional decision variable that will influence the sales y (a $n \times 1$ vector) in the n regions. Thus the model is of the type of a log linear production function as it is used in macro-economics.

For the optimal allocation problem in SRF models we need a target function that is suitable for sustainable growth rates.

Definition 1 (The stochastic allocation rule for ads expenditures).

We assume positive (and uncorrelated) sales $y_i, i = 1, \dots, n$ over n time periods and we assume that the total budget B_{tot} is allocated optimally over the n periods. The profit function to be maximized is $P = \sum_{i=1}^n d_i y(x_i)$. d_i is the marginal contribution of the product to the profit. Since we only considering 1 product, we can set $d_i = 1$. This leads to the following Lagrange function:

$$\sum_{i=1}^n d_i y(x_i) + \lambda (B_{tot} - \sum_{i=1}^n x_i)$$

The solution of this optimal allocation problem is given by setting the first derivative to zero, from where we find

$$d_i \partial y / \partial x_i \propto (y_x)_i = \lambda \quad \text{for } i = 1, \dots, n, \quad (3)$$

or that all derivatives of the sales $y(x_i)$ w.r.t. marketing effort x_i have to be constant. In cross-sections we refer to this allocation rule as 'Albers' rule because of Albers (1998).

This leads to the basic multiplicative (or Cobb-Douglas) type SRF model:

Definition 2 (The SRF model with latent partial derivatives). For observed regressor x , the multiplicative SRF(1) model $y = \beta_0 x^{\beta_1} e^\epsilon$ is defined as the following set of 2 log-normal densities:

$$\begin{aligned} \ln y &\sim N[\ln \beta_0 + \beta_1 \ln x, \sigma_y^2 I_n] \\ \ln y_x &\sim N[\log(\beta_0 \beta_1) + (\beta_1 - 1) \ln x, \sigma_y^2 I_n] \end{aligned} \quad (4)$$

where y_x is the first derivative of the SRF(1) model and with the parameters of the model given by $\theta = (\beta, \sigma_y^2)$.

Note that this is a non-linear model in β and also a very restricted model, since the sales observations y and the derivatives y_x follow normal distributions with the same variance. Furthermore, both equations are correlated and cannot be jointly estimated if y_x is unobserved. Thus, we need to look for better modeling strategies.

3.1 The SPD condition for SRF(1) models

We obtain an alternative SRF(1) model if we combine the assumptions for generating the first derivatives $\dot{y} = \ln y_x$ of the SRF model via the latent variable and the assumption of a stochastic optimal allocation (OA) rule like the stochastic Albers rule, like $\dot{y}_i = (\ln y_x)_i$

$$\dot{y}_i \mid \theta_\lambda \sim N[\lambda, \sigma_\lambda^2] \quad \text{for } i = 1, \dots, n \quad (5)$$

or $\ln y_x \sim N[\lambda 1_n, \sigma_\lambda^2 I_n]$. This means that the sales responses y and the decision variable x imply a prescription that marketing resources should be allocated according to the first derivative of the SRF model. Since the empirical observations across the n regions reveal some noise, we assume that the \dot{y}_i 's are independently normally distributed for given parameters $\theta_\lambda = (\lambda, \sigma_\lambda^2)$. These stochastic fluctuations of the derivatives across the n units are captured by the mean response λ , and the variance σ_λ^2 in assumption (5) imposes the looseness or strength of this target λ , the optimal behavior, from the actual

but not observed derivatives y_x . It measures in practice how good marketing people follow the prescription of the optimal allocation (OA) model.

If the \dot{y} 's could be observed, there would be no extra stochastic dependencies. In our model we have to proxy the unobserved derivatives by the realized derivatives of the multiplicative SRF function in (5):

$$\dot{y} = X\tilde{\beta} \quad \text{with} \quad \tilde{\beta} = \begin{pmatrix} \ln(\beta_0\beta_1) \\ \beta_1 - 1 \end{pmatrix} \quad (6)$$

Adding this stochastic partial derivative (SPD) constraint for the x regressor in the SRF model creates an behavioral model that the partial derivatives should be (approximately) equal across the regional units:

Lemma 1 (The SPD assumption for the SRF(1) model).

The combination of the stochastic optimal allocation (OA) rule (5) and the generation of the latent partial derivatives as in (4) implies the endogeneity of x in the SRF(1) model. Thus, the SPD assumption implies a normal distribution of the regressor x in the following way:

$$\ln x \mid \theta_\lambda \sim N[\mu_x(\theta_\lambda), \sigma_x^2(\theta_\lambda)], \quad (7)$$

with the parameters $\theta_\lambda = (\lambda, \sigma_\lambda^2)$ and mean and variance

$$\mu_x = \frac{\ln(\beta_0\beta_1) - \lambda}{1 - \beta_1}, \quad \sigma_x^2 = \frac{\sigma_\lambda^2}{(\beta_1 - 1)^2} \quad (8)$$

Proof. There are several ways to derive the result. One leads via the transformation rule for random variables to the Jacobian of $\ln x$ is just $1/|\beta_1 - 1|$. The other approach just equates equations (5) and (4) and solves for $\ln(x)$. One way to see how the SPD assumption translates to an assumption about the x is to write the exponent of the density (5) and use the log derivative (4)

$$\begin{aligned} (\ln y_x - \lambda)^2 / \sigma_\lambda^2 &= (\ln(\beta_0\beta_1) + (\beta_1 - 1)\ln x - \ln\lambda)^2 / \sigma_\lambda^2 = \\ &= \left(\frac{\log(\beta_0\beta_1) - \lambda}{1 - \beta_1} - \ln x \right)^2 (\beta_1 - 1)^2 / \sigma_\lambda^2 \\ &\propto p(\ln x \mid \mu_x, \sigma_x^2) \end{aligned}$$

Finally, we can define the SRF(1)-OA and the SRF(1)-SPD model in the following way:

Definition 3 (The SRF(1)-SPD model).

(a) *The SRF(1)-SPD model is based on the multiplicative SRF model $y = \beta_0 x^{\beta_1} e^\epsilon$, the endogeneity of x and the stochastic Albers rule (Definition 1), which result in the following set of 3 log-normal densities:*

$$\begin{aligned} \ln y \mid \theta_y &\sim N[\ln \beta_0 + \beta_1 \ln x, \sigma_y^2] \\ \Rightarrow \ln x \mid \theta_\lambda &\sim N[(\ln \beta_0 + \ln \beta_1 - \ln \lambda) / (\beta_1 - 1), \sigma_\lambda^2 / (\beta_1 - 1)^2] \\ \ln y_x \mid \theta_\lambda &\sim N[\lambda, \sigma_\lambda^2], \end{aligned} \quad (9)$$

where y_x is the first derivative of the SRF(1) model and the parameters of the model are given by $\theta = (\beta, \lambda, \sigma_y^2, \sigma_\lambda^2)$ and $\theta_\lambda = (\lambda, \sigma_\lambda^2)$, $\theta_y = (\beta, \sigma_y^2)$. The " \Rightarrow " denotes the derived distribution for $\ln x$, making the SPD and the constant allocation assumption.

(b) **The SRF(1)-OA model** that generates the endogeneity of x indirectly, and leads to the following reduced set of equations, with the restriction $\beta_1 > 0$

$$\begin{aligned} \ln y \mid \theta_y &\sim N[\ln \beta_0 + \beta_1 \ln x, \sigma_y^2] \\ \ln y_x \mid \theta_\lambda &\sim N[\lambda, \sigma_\lambda^2]. \end{aligned} \quad (10)$$

Again, $\ln y_x = X\tilde{\beta}$ is the realized derivative (6), and thus just a linear transformation of the regressors X and the β coefficients, and therefore for the control variable x , say $a + bx$. Therefore, this stochastic OA assumption implies implicitly a distribution for x .

For statistical inference we can estimate the parameter vector by maximum likelihood or by MCMC, assuming a prior density given by $p(\theta)$. In the next section we outline the MCMC procedure.

3.2 MCMC estimation in the SRF(1)-OA model

This section develops the MCMC estimation for the SRF(1)-OA model. The optimal allocation (OA) rule in the SRF model requires a first derivative, which can be not observed by data, and therefore has to be introduced into the model as a latent variable.

The latent variable defines another equation in the D/S system that can be generated conditionally through the assumptions of the system. The latent variable $\dot{y} = \ln y_x$ is considered to be a homolog (i.e. over-parameterized) parameter vector that is computed or estimated from the demand or y-equation. Finally, the observed data $\mathcal{D} = (\ln y, \ln x)$ and the latent variable $\ln y_x$ are modeled by the joint density $p(\ln y, \ln x, \ln y_x)$, which decomposes in the general case as

$$p(\ln y \mid \ln x, \beta, \sigma_x^2, \dots) p(\ln x \mid \beta, \ln y_x, \dots) p(\ln y_x \mid \lambda, \sigma_{\lambda^*}^2, \dots).$$

Because the realized derivative $\ln y_x = X\tilde{\beta}$ with $X = (1_n : x)$ is generated directly as a linear combination of the x variable, the density for $\ln x$ in (11) implies a likelihood function for x . Thus, the likelihood function for the SRF(1)-OA model is

$$l(\theta \mid \mathcal{D}) = N[\ln y \mid \mu_y, \sigma_y^2 I_n] N[\ln x \mid \mu_x, \sigma_\lambda^2 I_n] \quad (11)$$

with the conditional means

$$\mu_y = \ln \beta_0 + \beta_1 \ln x \text{ and } \mu_x = (\ln \beta_0 + \ln \beta_1 - \lambda) / (\beta_1 - 1). \quad (12)$$

The prior density for $\theta = (\beta, \sigma_y^{-2}, \sigma_\lambda^{-2})$ and $\ln y_x$ is

$$p(\theta) = N[\beta \mid \beta_*, H_*] N[\lambda \mid \lambda_*, \sigma_{\lambda_*}^2] N[\ln y_x \mid \lambda, \sigma_\lambda^2] \prod_{j \in \{y, \lambda\}} Ga[\sigma_j^{-2} \mid \sigma_{j_*}^2, n_{j_*}] \quad (13)$$

Thus, the SRF(1)-OA model consists of

1. The prior density (13),
2. The likelihood function (6),
3. The realized derivative (11).

The posterior distribution $p(\theta \mid \mathcal{D}) \propto l(\mathcal{D} \mid \theta)p(\theta)$ is simulated by MCMC.

Theorem 1 (MCMC in the SRF(1)-OA model).

The MCMC iteration in the SRF(1)-OA model with the likelihood function (11) and the prior density (13) takes the following draws of the full conditional distributions (fcd):

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_λ^{-2} from $\Gamma[\sigma_\lambda^{-2} \mid s_{\lambda^{**}}^2, n_{\lambda^{**}}]$*
4. *Draw λ from $N(\lambda \mid \lambda_{**}, s_{\lambda^{**}}^2)$*
5. *Compute $\dot{y} = \ln y_x$ from $N[\dot{y} \mid \mu_{\dot{y}}, (s_y^2)I_n]$*
6. *Draw $\beta = (\beta_0, \beta_1)$ from $p(\beta_0)$ and $p(\beta_1 \mid \beta_0)$*
7. *Repeat until convergence.*

The proof is given in the Appendix.

The marginal likelihood of model \mathcal{M} is computed by the Newton-Raftery formula

$$\hat{n}(\mathbf{y} \mid \mathcal{M})^{-1} = \frac{1}{n_{rep}} \sum_{j=1}^{n_{rep}} \left(\sum_{i=1}^n \ln l(\mathcal{D}_i \mid \mathcal{M}, \theta_j) \right)^{-1} l(\mathcal{D}_j \mid \mathcal{M}, \theta)^{-1} \quad (14)$$

where $\mathcal{D}_i = (\ln y_i, \ln x_i)$ is the i -th data observation and with the likelihood given in (11).

4 The AR-SRF(1)-OA model with SPD

In this section we describe the AR-SRF model because we want to demonstrate the effects of correlated time series for the estimation procedure. The AR-SRF(1) sales response function with one input variables x and the lagged endogenous variable y_{-1} is

$$\begin{aligned} y &= \beta_0 y_{-1}^\rho x^{\beta_1} e^\epsilon, \quad \text{or} \\ \ln y &= \rho \ln y_{-1} + \ln(\beta_0) + \beta_1 \ln x + \epsilon, \end{aligned} \quad (15)$$

where ϵ is a $N[0, \sigma_y^2]$ distributed error term. This leads to the reduced form equation

$$R \ln y = \ln y - \rho \ln y_{-1} = \ln(\beta_0) + \beta_1 \ln x + \epsilon$$

with

$$R = I_n - \rho L \quad \text{with} \quad L = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & \dots & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (16)$$

and L being a supra-diagonal or the AR(1) lag-shift matrix. By taking log's for the n observations we find for known ρ the reduced form regression model

$$R \ln y \sim N[\ln(\beta_0) + \beta_1 \ln x, \sigma_y^2 I_n]. \quad (17)$$

The mean of the reduced form regression is $\mu_y = X\beta$ with coefficients $\beta = (\ln \beta_0, \beta_1)$ and the regressor matrix $X = (1_n : \ln x)$ is given as before, where 1_n is a vector of 1's and x is the supply-side control variable that will influence the sales y (a $n \times 1$ vector). Thus, the model is again a log-linear production function as it is used in macro-economics. The stochastic partial derivative (SPD) assumption is applied to the reduced form equation and leads to the behavioral equation

$$\dot{g} = \partial R y / \partial x = \beta_0 \beta_1 x^{\beta_1 - 1} e^\epsilon.$$

The realized first derivative is

$$g_x = \frac{\partial R y}{\partial x} \Big|_x : n \times 1$$

evaluated at the vector x . This log (realized) derivative, for known β and SRF, is given as in the simple SRF model (18) by

$$\ln g_x = \mu_g + \epsilon = \ln(\beta_0 \beta_1) + (\beta_1 - 1) \ln x + \epsilon, \quad (18)$$

and therefore the log derivative $p(\ln g_x | \beta, x) = N[\mu_g, \sigma_y^2 I_n]$ with $\mu_g = \ln(\beta_0 \beta_1) + (\beta_1 - 1) \ln x$ is normally distributed.

This leads to the following AR-SRF(1)-OA (optimal allocation) model:

Definition 4 (The AR-gSRF(1) model with the partial derivative as latent variable). For observed y and x the the AR-gSRF(1) model (in reduced form) is defined with R as in (16) and $\beta_1 > 0$ as the following set of 2 log-normal densities:

$$\begin{aligned} R \ln y &\sim N[\ln \beta_0 + \beta_1 \ln x, \sigma_y^2 I_n] \\ \ln y_x &\sim N[\ln(\beta_0 \beta_1) + (\beta_1 - 1) \ln x, \sigma_y^2 I_n], \end{aligned} \quad (19)$$

where y_x is the first derivative of the AR-gRF(1)-OA model and $\theta = (\beta, \rho, \sigma_y^2)$ are the parameters of the model.

4.1 A loss function for growth rates

We consider the growth factor over time and we argue that the growth factors are the target of the SRF models to monitor long term sales growth. The growth rates are obtained from the growth factors by taking logs.

The total budget available is B_{tot} and instead of maximizing the profit directly, we look for a function that maximizes the sustainable growth of sales. Thus, the criterion to be maximized is slightly different:

$$Q = \sum_{i=1}^n g_t(x_i) \quad \text{where} \quad g_t(x_i) = \frac{y_t(x_i)}{y_{t-1}(x_i)}.$$

$g_t(x_i)$ denotes the growth factor of the sales: The growth factors are needed to ensure positive vales of the SRF model. This function is correlated with the profit function in (1). As a side constraint we assume that the company is interested in a sustainable growth path, which is expressed as deviation between the average growth rate

$$\bar{g} = \frac{1}{n} \sum_{i=1}^n g_t \tag{20}$$

from a target growth rate g_* over n periods. These considerations lead to the following optimisation problem using the Lagrange function for the growth factors of sales, which mimics the Albers rule (1) for a time period of length n :

Definition 5 (An optimal allocation (AO) rule for sustainable growth).

We consider the growth factors $g_t(x_i)$ of sales that depend on the ads variable x for n periods and we assume that the ads expenditures are allocated according to a sustainable growth path as in (20)

$$G(x_i, \lambda) = \sum_{i=1}^n g_t(x_i) + \lambda(n g_* - \sum_{i=1}^n g_t(x_i)). \tag{21}$$

The solution of this Lagrange problem requires setting the first derivative $\partial G / \partial x_i$ to zero, from where we find

$$g_x = \partial g_t / \partial x_i = \lambda \quad \text{for} \quad i = 1, \dots, n, \tag{22}$$

or that all derivatives of the sales growth factors $g_t(x_i)$ w.r.t. marketing effort x_i have to be constant.

While the Albers rule is applicable for sales in n regions, the sustainable growth rule works for time series. It is set up in a similar way so that we have a simple ads allocation rule over time. Note the similarity to the original stochastic Albers rule. If a long-term planing horizon and the budget becomes important, then the cross-sectional units are replaced by time series data. Thus, the ads budget can be allocated in a simple way over the planing period.

4.2 MCMC estimation in the AR-gSRF(1)-OA model

This section shows the MCMC estimation of the AR-gSRF(1)-OA model. The likelihood function is a function of the observed data and the parameters $\theta = (\beta, \sigma_y^{-2}, \lambda, \rho)$, where the matrix R involves the correlation coefficient ρ

$$l(\theta | \mathcal{D}) = N[R \ln y | \mu_y, \sigma_y^2 I_n] N[\ln y_x | \lambda, \sigma_{\lambda^*}^2 I_n] \quad (23)$$

with the conditional means μ_y and μ_x given by

$$\mu_y = \ln \beta_0 + \beta_1 \ln x \quad \text{and} \quad \mu_x = (\ln \beta_0 + \ln \beta_1 - \lambda) / (\beta_1 - 1). \quad (24)$$

Because the realized derivative $\ln y_x = X\tilde{\beta}$ with $X = (1_n : x)$ is generated directly as a linear combination of the x variable, the second density in (23) translates actually to a likelihood function for x . The prior density for θ is

$$p(\theta) = N[\beta | \beta_*, H_*] N[\lambda | \lambda_*, \sigma_{\lambda^*}^2] \prod_{j \in \{y, \lambda\}} Ga[\sigma_j^{-2} | \sigma_{j^*}^2, n_{j^*}] \quad (25)$$

where all the parameter with a '*' index denote known hyper-parameters of the prior distribution. Finally, the AR-gSRF(1)-OA model consists of

1. The prior density (25),
2. The likelihood function (23),
3. The realized derivative (6).

The posterior distribution $p(\theta | \mathcal{D}) \propto l(\mathcal{D} | \theta)p(\theta)$ is simulated by MCMC.

Theorem 2 (MCMC in the AR-gSRF(1)-OA model).

The MCMC iteration in the AR-gSRF(1)-OA model with the likelihood function (23) and the prior density (25) takes the following draws of the full conditional distributions (fcd):

1. *Starting values: set $\beta = \beta_{OLS}$ and $\lambda = 0$*
2. *Draw σ_y^{-2} from $\Gamma[\sigma_y^{-2} | s_{y^{**}}^2, n_{y^{**}}]$*
3. *Draw λ from $N(\lambda | \lambda_{**}, s_{\lambda^{**}}^2)$*
4. *Compute $\dot{g} = \ln g_x = X\beta$*
5. *Draw $\beta = (\beta_0, \beta_1)$ from $p(\beta_0)$ and $p(\beta_1 | \beta_0)$*
6. *Draw ρ by a gridgy Gibbs step using $p(\rho | \mathcal{D}, \dots)$.*
7. *Repeat until convergence.*

Proof. The proof is almost identical to Theorem 1, except that we need one more fcd for the extra parameter ρ . Furthermore only the fcd's for the first layer for the log-y equation, β and the residual variance are affected by the reduced form transformation $y \rightarrow Ry$: The residuals in (33) change to $e_y = R \ln y - X\beta$ and the fcd (29) for β changes to

$$p(\beta | \mathcal{D}, \dots) \propto N[\beta | \beta_*, H_*] N[R \ln y | X\beta, \sigma_y^2 I_n] N[\ln x | \mu_x, \sigma_{\lambda^*}^2 / (1 - \beta_1)^2 I_n] \quad (26)$$

and we have to make the variable change in (30) to

$$b_{\#} = H_{\#} [H_{*}^{-1}b_{*} + \sigma^{-2}X'Ry].$$

For the fcd of the correlation coefficient ρ we need as proposal density the normal distribution based on the OLS estimate. The fcd is proportional to

$$p(\rho) = p(\rho | \mathcal{D}) \propto \sigma_y^{-2T} \exp \left[-\frac{1}{2\sigma^2} (Ry - I^{\dagger}\tau)'(Ry - I^{\dagger}\tau) \right].$$

The griddy Gibbs step: We evaluate a grid of 100 rho points around the MLE of the approximate linear model: $y = \rho Ly + \epsilon$. Under normality we have $\rho \sim N[\hat{\rho}, \sigma_{\rho}^2]$ with $\hat{\rho} = y'Ly/y'L'Ly$ and $\sigma_{\rho}^2 = \sigma_y^2/y'L'Ly$. The exponent is $(y - \rho y_{-1})'(y - \rho y_{-1})/\sigma^2 = (\rho - \hat{\rho})^2 S_y/\sigma_y^2 \propto N[\rho | \hat{\rho}, \sigma_y^2/S_y]$.

Note that the MCMC algorithm for the AR-gSRF-X model parallels the structure in Theorem 2, only the variables in the regressor matrix have to be arranged as $X = (1 : z : x)$, in blocks of exogenous and endogenous variable where z is the exogenous variable.

Model choice: The marginal likelihood of model \mathcal{M} is computed by the Newton and Raftery (1994) formula (14) with the likelihood given in (23).

5 Example: Ads and beer consumption in Germany

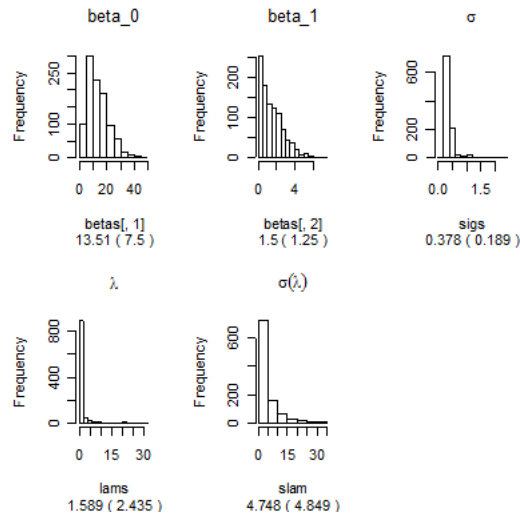
The German beer sales in hekto-liter (hl) and the marketing expenditures (in mio Euros) for 1997-2010 are found in Table 1. (Source: The German Statistische Bundesamt)

Table 1: German beer and ads data 1997-2010

Year	Sales (hl)	Marketing
1997	103	402,00
1998	100,18	431,00
1999	110,10	380,00
2000	109,80	388,00
2001	107,80	360,00
2002	107,80	347,00
2003	105,60	331,00
2004	105,90	364,00
2005	105,40	410,00
2006	106,80	374,70
2007	104,00	399,30
2008	102,90	401,80
2009	100,00	350,30
2010	98,30	376,88

The MCMC densities of the parameters of the SRF(1)-OA model are in Figure 1.

Fig. 1. Posterior MCMC betas of the SRF(1)-OA model, 1998-2000



The connection between beer and wine consumption is quite strong, as we see in Figure 5, but there is also a surprising negative correlation between the market share of wine and beer ads (in Euros) in Germany.

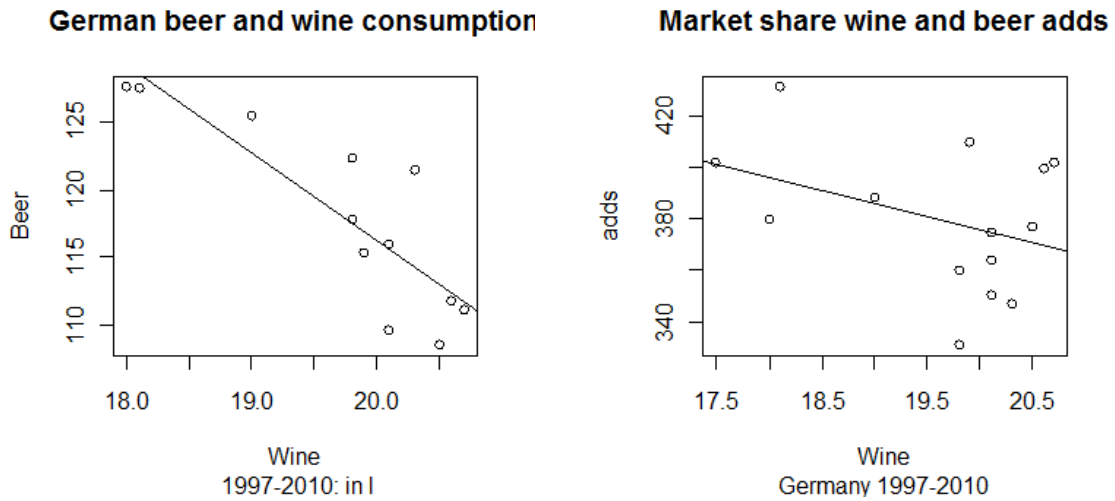
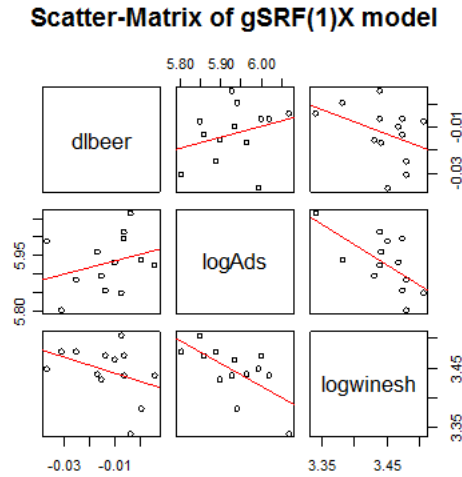


Fig. 2. a) Beer & wine cons. 1998-2000; b) Market share of wine and beer ads

5.1 Results for the AR-SRFX model

The log data of the SRF(1)X(1)-AR(1) model are displayed by a scatter-plot matrix with bivariate regression lines in Figure 3.

Fig. 3. The log data for SRFX-AR model in Germany 1999-2010



The modeling strategy is as follows. We start with the most complex model as its MCMC estimation is given for the AR-SRFX-AO model in Theorem 2. Also, because of negative MCMC diagnostics, we dropped the assumptions of a model imposing a SPD prior for the control variable x and we prefer to use the SRFX-AO model. Furthermore, we prefer for the σ_λ^{-2} parameter to be fixed at a tight constant, implying that the (stochastic) sustainable growth allocation rule is taking place in a rather tight narrow band. Based on the OLS estimate of ρ , the σ_λ^{-2} can be easily fixed at the variance of the latent variable $\ln y = y_x$ evaluated at the OLS β coefficients. The MCMC estimates of the AR-SRF model are:

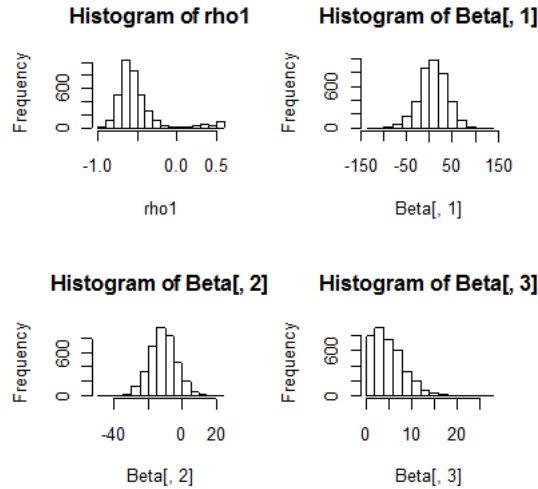
	mean beta	SD beta
beta[1]	7.709274	30.199488
beta[2]	-11.405636	8.275701
beta[3]	5.009594	3.499446

and for the auto-correlation coefficient we find as average over the MCMC sample $\rho = -0.5089122, SD(\rho) = 0.2873213$.

The density estimates can be seen in Figure 4. Convergence was achieved very fast and the use of the gridy Gibbs method created no autocorrelation in the ρ -runs. (A Metropolis-Hastings algorithm did not work so well.) The range of coefficients is rather wide, but this is not surprising since there are only 12 observations. (Classical results based on asymptotic distributions would not

work well for this data set.) The elasticity on log ads (the endogenous variable) has to be positive which is the case in 75% of the number of iterations. Interestingly, by discarding the negative draws, we get about the same type of distributions (histogram shapes) of the coefficients.

Fig. 4. The density estimates for SRFX-AR model in Germany 1999-2010



Following our "general to more simpler" specification philosophy we find, that fixing certain hyper-parameters at a reasonable value yield better estimation results for the coefficients of the SRF model in the first stage. Because there are many parameters to estimate plus a latent variable, the Bayesian analysis improves if there are fewer 'free' parameters to estimate.

6 Conclusions

The paper has shown that the class of sales response models is large and flexible enough to cope with shrinking sales in sales models with advertisement expenditures. The results of the sales growth response function (gSRF) model point in the right direction, but only if important possible marketing behaviors have been appropriately implemented in the model.

As an additional consideration for the estimation of an SRF model in a time series context we have proposed a specification that allows for AR(1) errors. With this assumption we leave the framework of easy simulations in the MCMC algorithm using normal and gamma distributions. We found that the use of the gridly Gibbs sampler for the autocorrelation coefficients leads

to a quick mixing of the sampler without the perils of long autocorrelation when a Metropolis-Hastings step is used.

Given the large variety of models (SRF-X and AR-SRF-X) we have tried and from the non-significant estimation results we conclude that the beer industry has not reacted in a proper way to fight the shrinking sales in the beer market in Germany over the last decade. For future work, many further extensions of this flexible class of SRF-X models are possible. First of all there is the question of the right or appropriate functional form of the SRF-X class models. Following the suggestions of Kao et al. (2005) more research is needed, especially if time series and shrinking markets have to be considered. Secondly, there is the question of the appropriate marketing actions and strategies undertaken by the supply side of the market.

A possible next step for a better model choice is the use of Bayesian model averaging (BMA) techniques. Thus, there is room for more theory as how market participants react either as consumers to marketing efforts or as marketing strategists who react either to sales developments or company policies. System estimation would be required if more than 1 marketing channels should be optimized or if marketing efforts depend also on the sales performance of the competitors.

References

1. Albers S. (1998) Regeln fuer die Allokation eines Marketingbudgets auf Produkte oder Marktsegmente, *Zeitschrift fuer Betriebswirtschaftliche Forschung* 50, 211-229.
2. Anselin L.(1988) Spatial Econometrics. In: B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*. Blackwell Publishing Ltd AD, 310-330.
3. Baier, D. and Polasek W. (2010), Marketing and Regional Sales: Evaluation of Expenditure Strategies by Spatial Sales Response Functions, *in: Studies in Classification, Data Analysis, and Knowledge Organization, Vol. 40, 673-682*. .
4. Kao L.-J., C.-C. Chiu, T.J. Gilbride, T. Otter, and G.M. Allenby (2005) Evaluating the Effectiveness of Marketing Expenditures. *Working Paper*, Ohio State University, Fisher College of Business.
5. Newton, M.A. and A. E. Raftery (1994) Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B*, 56, 3-48
6. Polasek W. (2010a), Sales Response Functions (SRF) with Stochastic Derivative Constraints, *Institute fuer hoehere Studien, Wien*.
7. Polasek W. (2010b), Endogeneity and Exogeneity in Sales Response Functions, *to appear in GFKL 2010* .
8. Polasek W. (2011), Multi-level panel models for regional beer sales in Germany, *Institute fuer hoehere Studien, Wien*.
9. Christian Ritter and Martin A. Tanner (1992), Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association, Vol. 87, No. 419, 861- 868*
10. Rossi P.E., G.M. Allenby, and R. McCulloch (2005) *Bayesian Statistics and Marketing*. John Wiley and Sons, New York.

7 Appendix

7.1 Proof of Theorem 1 (MCMC in the SRF-OA model)

Proof. The full conditional densities (fcd's) are as follows:

1. The fcd for λ , the average utility level can be estimated in the same way as before:

$$p(\lambda \mid \mathcal{D}, \dots) \propto N[\lambda \mid \lambda_*, s_{\lambda_*}^2] N[\ln y_x \mid \lambda 1_n, \sigma_\lambda^2 I_n] \propto N[\lambda \mid \lambda_{**}, s_{\lambda_{**}}^2] \quad (27)$$

with $s_{\lambda_{**}}^{-2} = s_{\lambda_*}^{-2} + n\sigma_\lambda^{-2}$ and from (5) we find in the exponent the quadratic form $(\ln y_x - \lambda 1_n)' \sigma_\lambda^{-2} (\ln y_x - \lambda 1_n)$

$$\lambda_{**} = s_{\lambda_{**}}^2 (s_{\lambda_*}^{-2} \lambda_* + n\sigma_\lambda^{-2} 1_n' (\ln y_x)),$$

where σ_λ^2 is the variance of y_x and the realized y_x 's are evaluated at the current β .

2. The fcd for $z = \ln y_x$ under SPD is

$$\begin{aligned} p(\ln y_x \mid \mathcal{D}, \dots) &\propto N(\ln y_x \mid \mu_z, \sigma_z^2 I_n) N[\ln y_x \mid \lambda 1_n, \sigma_\lambda^2 I_n] \\ &= N[z \mid \mu_{z^{**}}, s_{z^{**}}^2 I_n]. \end{aligned} \quad (28)$$

with $s_{z^{**}}^{-2} = \sigma_z^{-2} + \sigma_\lambda^{-2}$ and $\mu_{z^{**}} = s_{z^{**}}^{-2} (s_z^{-2} \mu_z + \sigma_\lambda^{-2} 1_n \lambda)$.

3. The fcd for β coefficients is

$$p(\beta \mid \mathcal{D}, \dots) \propto N[\beta \mid \beta_*, H_*] N[\ln y \mid X\beta, \sigma_y^2 I_n] N[\ln y_x \mid \lambda 1_n, \sigma_\lambda^2 I_n] \quad (29)$$

since the third density of $\ln x$ in (29) contains the β coefficients in a non-linear way and μ_x is given in (12). To avoid a Metropolis step we get an analytical solution by combining the 3 components of normal densities in 3 steps.

Step 1: The first two normal densities can be combined in the usual way to

$$\begin{aligned} N \left[\beta \mid \beta_{\#} = \begin{pmatrix} \beta_{0\#} \\ \beta_{1\#} \end{pmatrix}, H_{\#} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix} \right] &\text{ with} \\ H_{\#}^{-1} &= H_*^{-1} + \sigma^{-2} X'X, \\ b_{\#} &= H_{\#} [H_*^{-1} b_* + \sigma^{-2} X'y], \end{aligned} \quad (30)$$

where the index '#' indicates an auxiliary result.

Step 2a: The conditional bivariate normal density in (30) for $\beta_1 \mid \beta_0$ is:

$$\begin{aligned} p(\beta_1 \mid \beta_0) &= N[\beta_{1.0}, \sigma_{1.0}^2] \text{ with} \\ \sigma_{1.0}^2 &= h_{11} - h_{10}h_{01}/h_{00} = h_{11}(1 - \rho_{01}^2), \\ \beta_{1.0} &= \beta_{1\#} + h_{10}(\beta_0 - \beta_{0\#})/h_{00}. \end{aligned} \quad (31)$$

Note that ρ_{01}^2 is the squared correlation coefficient, defined as $\rho_{01}^2 = \frac{h_{10}^2}{h_{00}h_{11}}$.
 Step 2b: The general case for the conditional normal density in (30) for $\beta_1 | \beta_0$:

$$\begin{aligned} p(\beta_1 | \beta_0) &= N[\beta_{1.0}, \sigma_{1.0}^2] \quad \text{with} \\ \sigma_{1.0}^2 &= h_{11} - h_{10}h_{00}^{-1}h_{01} \\ \beta_{1.0} &= \beta_{1\#} + h_{10}h_{00}^{-1}(\beta_0 - \beta_{0\#}). \end{aligned} \quad (32)$$

The variables in the SRF regression model need to be ordered in such a way that the component with '0' contains the intercept (and the z variables of the SRFX model), while the component with '1' contains the endogenous variable x .

Step 3: Simulate the positive β_1 coefficient either by keeping only those draws that are positive or draw from a truncated normal density restricted to the positive real line. The third density in (29) is also restricting the draws and follows the same drawing approach using the conditional normal density. The following Metropolis-Hastings step is used: We use a random walk chain for the proposal β^{new}

$$\beta^{new} = \beta^{old} + \mathcal{N}[0, c_\beta I_k],$$

where k is the dimension of β . c_β is a tuning constant for the variance of the proposal. The acceptance probability involves the posterior fcd density $p(\beta = p(\beta | \mathcal{D}, \dots))$ in (29) and is given by

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

where we accept only proposals with $|\beta_1^{new}| > 0$

4. The fcd for σ_y^{-2}

$$p(\sigma_y^{-2} | \mathcal{D}, \dots) \propto Ga[\sigma_y^{-2} | \sigma_{y^{**}}^2 n_{y^{**}}/2, n_{y^{**}}/2] \quad (33)$$

with $n_{y^{**}} = n_{y^*} + n$ and $n_{y^{**}}\sigma_{y^{**}}^2 = n_{y^*}\sigma_{y^*}^2 + e_y'e_y$, where $e_y = \ln y - X\beta$ being the current residuals of the log-y equation.

5. Only in case where the stochastic OA variance σ_λ^{-2} will be estimated: The fcd for σ_λ^{-2}

$$p(\sigma_\lambda^{-2} | \mathcal{D}, \dots) \propto Ga[\sigma_\lambda^{-2} | \sigma_{\lambda^{**}}^2, n_{\lambda^{**}}] \quad (34)$$

with $n_{\lambda^{**}} = n_{\lambda^*} + n$ and $n_{\lambda^{**}}\sigma_{\lambda^{**}}^2 = n_{\lambda^*}\sigma_{\lambda^*}^2 + e_x'e_x + e_\lambda'e_\lambda$ and the residuals $e_x = \ln x - \mu_x$ and $e_\lambda = \ln y_x - \lambda 1_n$ (or $e_\lambda'e_\lambda = \sum_i (\ln y_{x,i} - \lambda)^2$). This is because we have 2 variance sources

$$\begin{aligned} p(\sigma_\lambda^{-2} | \mathcal{D}, \dots) &\propto \left(\frac{\sigma_\lambda^2}{(1 - \beta_1)^2}\right)^{-n/2} \exp\left[-\frac{1}{\sigma_\lambda^2}(\ln x - \mu_x)'(\ln x - \mu_x)(1 - \beta_1)^2\right] \\ &(\sigma_\lambda^2)^{-n/2} \exp\left[-\frac{1}{\sigma_\lambda^2}(\ln y_x - \lambda 1_n)'(\ln y_x - \lambda 1_n)\right] \end{aligned}$$

7.2 The griddy Gibbs sampler

This procedure was described in Ritter and Tanner (1992). Consider a m -dimensional posterior density $p(\theta_1, \dots, \theta_m)$ that is estimated via MCMC and where the conditional distribution $p(\theta_i | \theta_j, j \neq i)$ is untractable but univariate. If it is difficult to directly sample from $p(\theta_i | \theta_j, j \neq i)$, the idea is to form a simple approximation to the inverse cdf based on the evaluation of $p(\theta_i | \theta_j, j \neq i)$ on a grid of points. This leads to the following 3 steps:

Step 1. Evaluate $p(\theta_i | \theta_j, j \neq i)$ at $\theta_i = x_1, x_2, \dots$ to obtain w_1, w_2, \dots, w_n .

Step 2. Use w_1, w_2, \dots, w_n to obtain an approximation to the inverse cdf of $p(\theta_i | \theta_j, j \neq i)$.

Step 3. Sample a uniform $U(0, 1)$ deviate and transform the observation via the approximate inverse cdf.

Remark 1: The function $p(\theta_i | \theta_j, j \neq i)$ need be known only up to a proportionality constant, because the normalization can be obtained directly from the w_1, w_2, \dots, w_n .

Remark 2: The grid x_1, x_2, \dots, x_n need not be uniformly spaced. In fact, good grids put more points in neighborhoods of high mass and fewer points in neighborhoods of low mass. One approach to address this goal is to construct the grid so that the mass under the current approximation to the conditional distribution between successive grid points is approximately constant.