IMPOSING CURVATURE AND MONOTONICITY RESTRICTIONS ON FLEXIBLE FUNCTIONAL FORMS

Hendrik Wolff, Thomas Heckelei and Ron C. Mittelhammer

January 16, 2004

Abstract

In many areas of economic analysis, economic theory restricts the shape as well as other characteristics of functions used to represent economic constructs. Obvious examples are the monotonicity and curvature conditions that apply to utility, profit, and cost functions. Commonly, these regularity conditions are imposed either *locally* or *globally*. Here we extend and improve upon currently available estimation methods for imposing regularity conditions by imposing regularity on a connected subset of the regressor space. This method offers important advantages over the local approach by imposing theoretical consistency not only locally, at a given evaluation point but also within the whole empirically relevant region of the domain associated with the function being estimated. The method also provides benefits relative to the global approach, through higher flexibility, which generally leads to a better model fit to the sample data compared to the global imposition of regularity.

Specific contributions of this paper are (a) to increase the computational speed and tractability of imposing regularity conditions in estimation, (b) to provide regularity preserving point estimates, (c) to avoid biases existent in previous applications, and (d) to illustrate the benefits of the regional approach via numerical simulation results.

Keywords: Nonlinear Inequality Constraints, Flexible Functional Forms, Metropolis-Hastings, Accept-Reject Algorithm, Cost Function, Regularity Conditions

JEL Code: C51 - Model Construction and Estimation,
D21 - Firm Behavior
C11 - Bayesian Analysis

Thanks are due to Guido Imbens, Haijun Li, Siegfried Schaible, Mark Voornefeld and Arnold Zellner for correspondence and valuable discussions during the course of the research. The first author is grateful to the IMPACT-Center, Washington State University, and the German Academic Exchange Service (DAAD) for financial support during his 12 months study period in the United States in 2002 and 2003. Contact: Hendrik Wolff: wolff@are.berkeley.edu, Thomas Heckelei: heckelei@agp.uni-bonn.de and Ron C. Mittelhammer: mittelha@wsu.edu.

1. Introduction

In many areas of economic analysis regularity conditions, derived by economic theory, restrict the shape of the mathematical functions used to model technology and/or economic behavior. Examples are curvature and monotonicity restrictions which apply to indirect utility, expenditure, production, profit, and cost functions. During the last thirty years it has become standard to use second-order flexible functional forms for empirical analyses, such as the Translog and the Generalized Leontief, which have the ability to attain arbitrary local elasticities, i.e., at one point in the regressor space. More recently, higher (than second) order series expansions, such as the Fourier and the Asymptotically Ideal Production Model (AIM), have been suggested. These representations promise a better fit to the data as they transition from local to global flexibility and as the order of expansion increases.

Unfortunately, these estimated functions that model economic behavior frequently violate curvature and monotonicity restrictions and the propensity for such violations increases with the order of flexibility. Violations can lead to ambiguous forecasts and errant conclusions about economic behavior. Concerns related to the imposition of regularity conditions is as old as the literature on flexible functional forms and represents '*one of the most vexing problems applied economists have encountered*' DIEWERT and WALES (1987).

In this paper we propose and illustrate a Bayesian estimation procedure for imposing regularity conditions on a connected subset of the regressor space representing what we refer to as the *empirically relevant region*. We label this method as the *regional approach* to imposing regularity on a given parametric[1] function being estimated. It offers important advantages over the *local* approach by imposing theoretical consistency not only locally at a given evaluation point, but also over the whole empirically relevant region of the domain associated with the function being estimated. The method also provides benefits relative to the *global* approach, through higher

[1] To maintain shape restrictions in the nonparametric estimation framework, see MATZKIN, 1994. Instead, here we consider the estimation of a given parametric functional form subject to nonlinear inequality constraints.

flexibility derived from being less constraining, which generally leads to a better model fit to the sample data compared to the *global* imposition of regularity. In order to better understand the differences between the *regional*, *local* and *global* approach, we begin by discussing how previous methods handled the imposition of regularity.

1.1. The global approach

A widely applied partial solution to the problem of imposing regularity conditions is to devise parametric restrictions that impose the curvature conditions *globally*, i.e. at all values of the regressor space (see DIEWERT and WALES, 1987). For most[2] flexible functional forms, however, such restrictions come at the cost of limiting the flexibility of the functional form with regard to representing other economic relationships. For example, under the imposition of global concavity, the Generalized Leontief cost function does not allow for complementary relationships among inputs.

As recently noted by BARNETT (2002) and BARNETT and PASUPATHY (2003), the 'monotonicity' regularity condition has been mostly disregarded in estimation, leading to questionable interpretability of the resultant empirical economic models. A fundamental difficulty, however, is that imposing both curvature and monotonicity can extirpate the property of second order flexibility: For the special case of finite linear-in-the-parameters functional forms, which is the most common in empirical applications, LAU (1986:pp.1552-57) proved that flexibility is incompatible with global regularity if both concavity and monotonicity are imposed. Thus, maintaining higher order flexibility requires giving up *global regularity* (although one might

[2] An exception is the class of quadratic functional forms, e.g. the Generalized and Symmetric McFadden, on which the curvature is easily imposed on the parameters of the Hessian without destroying the flexibility property, as shown by LAU 1978 and DIEWERT and WALES (1987). However, if one wishes to impose curvature *and* monotonicity on functional forms, then the restrictions are functions of the parameters *and* the regressor variables. A solution to this problem is the purpose of this paper.

maintain *local flexibility*), which is a fact that does not seem to be generally appreciated in the literature on globally flexible functional forms.[3]

1.2. The local approach

The *local approach* maintains the flexibility property of a functional form if the regularity conditions are imposed at one selected point of the regressor space (i.e RYAN and WALES, 1998). The risk with this approach is that regularity may be violated in a neighborhood of this selected point. Because of this dilemma, the literature on flexible functional forms is characterized by a continual investigation for new functional forms that produce relatively large regular regions. Nonetheless, for a given data set, searching for alternate forms and applying and testing the regularity conditions on a case by case basis becomes an arduous task,[4] that can also be rife with statistical testing/verification problems. In 1984, GALLANT and GOLUB proposed an inequality-constrained optimization program to impose regularity conditions *locally* at each observed regressor value. Compared with the global approach, this method generally increases the fit of the model to the data. However, two problems remain: (a) the procedure becomes numerically difficult for large sample sizes and/or complicated constraints and (b) it is possible that the estimated form is irregular at points other than the sample observations. Hence, more general methods of imposing the regularity conditions are desirable and those which appear to be the most promising are summarized below in section 1.3.

---

[3] For example, a globally consistent second order Translog reduces the feasible parameter values of its squared terms to be zero, thus restricting the functional form to its (*second order inflexible*) first order series expansion, the Cobb-Douglas, which has constant elasticities.

[4] Examples of functional forms investigated are the Minflex Laurent (BARNETT 1985), Extended Generalized Cobb Douglas (MAGNUS 1979), Symmetric Generalized McFadden and Symmetric Generalized Barnett (DIEWERT and WALES 1987). Furthermore see the cited literature in BARNETT, GEWEKE and WOLFE (1991:p.10) and more recently DRISCOLL and BOISVERT (1991), TERRELL (1995, 1996), IVALDI ET AL. (1996), FLEISSIG, KASTENS and TERRELL (1997, 2000), JENSEN (1997), RYAN and WALES (1998), FISCHER, FLEISSIG and SERLETIS (2001) for studies evaluating these mentioned and other competing forms. We recommend BARNETT, GEWEKE and WOLFE (1991: pp.3-15) for an extensive and insightful review on the various developments, trials and errors in the history of using flexible functional forms.

1.3. Towards regional regularity

In order to circumvent the problem of the estimated form being irregular at points other than the sample observations, GALLANT and GOLUB discussed the possibility of imposing regularity conditions on a predefined regular region ψ of the regressor space by outlining a double inequality constrained optimization procedure. This *regional regularity approach* has the advantage that flexibility of the functional form can be maintained to a large degree while remaining theoretically consistent in the region where inferences will be drawn. In addition, imposing regional regularity generally leads to better forecasts than global regularity. However, GALLANT and GOLUB did not demonstrate the tractability of this approach and it seems that empirical implementation can be formidable with the currently available optimization tools.

It was not until 1996 that TERRELL advanced ideas relating to the empirical application of regional regularity. Instead of explicitly using a constrained optimization algorithm he decomposed the problem into a series of steps: First, a convex set ψ of the domain of the function is approximated by a dense grid consisting of thousands of singular regressor values. Second, using a Bayesian framework, an unconstrained posterior distribution of the parameter vector β, conditional on the endogenous variable $\mathbf{y}$, $p_u(\beta|\mathbf{y})$, is derived that does not incorporate the regularity conditions. Third, a Gibbs sampler is used to draw parameter vector outcomes from $p_u(\beta|\mathbf{y})$, and an Accept-Reject algorithm is applied to assess regularity for each outcome at all grid points. Finally, point estimates are derived and inferences are drawn based on the set of regular parameter vectors and its truncated posterior distribution. This procedure has two problems: (a) Due to the approximation of the relevant regressor space by the grid, the possibility that the function is irregular for some non-grid points cannot be eliminated. In this sense TERRELL does not impose regional regularity (on a connected set) but he imposes local regularity at multiple singular points. (b) The Gibbs simulator requires sampling from the entire support Θ of the unconstrained posterior $p_u(\beta|\mathbf{y})$. However, this can be time consuming if, as is often the case in practice, the regular region is only a small subset of Θ (TERRELL 1996).

To overcome the latter problem, GRIFFITHS, O'DONNELL and TAN CRUZ (2000:p.116) suggested using a Metropolis-Hastings Accept-Reject Algorithm (subsequently denoted as MHARA). Compared to the Gibbs algorithm, MHARA may increase the probability that sampled parameter vectors are regular, and therefore may be faster than Gibbs sampling. However, the related literature on MHARA[5] did not pursue the regional approach further, but rather continued to impose local regularity without proving the theoretical consistency on the domain of interest.

1.4. Objectives and organization

The principle goal of this paper is to improve upon current methods of imposing regularity conditions. Improvement is achieved by pursuing the following two objectives with regard to estimated functions:

(I) economic theory is not violated on a connected subset ψ which encompasses the empirically relevant region of the regressor space, and

(II) for a given function, the model fit – as judged by *any specified* scalar measure of fit on the regular parameter space – is optimized.

We promote the application of regional regularity by combining elements of TERRELL's Bayesian approach with the MHARA. This defines an alternative methodology that substantially mitigates previous difficulties and inconsistencies in applying the regional regularity concept. New features of our proposed method include:

1. a set of sufficient conditions for which regularity is guaranteed at 'any' point in ψ (objective I). If these conditions are satisfied, a twofold benefit results:
    i) Imposition of regularity in ψ does not rely on a grid approximation, and
    ii) the computational speed of the Accept-Reject algorithm is greatly enhanced as only a few critical points need to be checked for regularity.

[5] Literature on applications of MHARA include KOOP, OSIEWALSKI and STEEL (1994), O'DONNELL, SHUMWAY and BALL (1999), GRIFFITHS, O'DONNELL and TAN CRUZ (2000), GRIFFITHS (2001, 2003), CHUA, GRIFFITHS and O'DONNELL (2001), CUESTA ET AL. (2001), KLEIT and TERRELL (2001), O'DONNELL, RAMBALDI and DORAN (2001) and O'DONNELL and COELLI (2003).

2. allowing ψ to be some connected non-convex set, which can significantly increase the model fit achievable from estimation (objective II).
3. demonstrating that the commonly used MHARA sampling technique suffers from an upward bias of posterior density values in the neighborhood of the truncation boundary. We provide a simple bias-mitigating alternative.
4. demonstrating that the commonly used posterior mean may be inappropriate as a point estimate of model parameters due to the potential violation of regularity conditions. As an alternative, we suggest two regularity-preserving point estimates:
    i) the posterior mode
    ii) the parameter vector that minimizes error loss subject to regularity constraints.

The organization of the paper is as follows: In section 2, we motivate the methodology and outline the estimation procedure in general terms. Section 3 provides a more technical description of procedures and discusses the four methodological contributions. Examples using the AIM and Translog functional forms are given in section 4 in order to illustrate the methodology and demonstrate empirical relevance. A final section concludes and the appendix contains proofs of new propositions.

2. Methodological background

This section provides a general overview of the regularity conditions to be imposed, the Bayesian context of the problem, the Markov Chain Monte Carlo (MCMC) algorithm used, and the Accept-Reject Algorithm.

2.1. The cost function example

For illustrative purposes, consider estimating a system of input demand equations imposing a regular region on the underlying unit cost function, $f(\mathbf{p};\beta)$, whereby $\mathbf{p} = [p_1, p_2,\ldots, p_K]^T \in \pi$ are K input prices, π denotes the orthant of strictly positive prices in $\mathbb{R}^K$, and $\beta \in \Theta$ is the parameter vector to be estimated. According to economic theory $f(\mathbf{p};\beta)$ must be concave and

nondecreasing in $\mathbf{p}$ (MAS-COLELL, WHINSTON and GREEN, 1995:p.141). The regularity conditions to be imposed on a subset $\boldsymbol{\psi}$ of the price space $\boldsymbol{\pi}$ can be characterized by H elementary *Inequality Constraint Functions* (ICFs), $\mathbf{i} \equiv [i_1, i_2, \ldots, i_H]$: $(\boldsymbol{\pi} \times \boldsymbol{\Theta}) \rightarrow \mathbb{R}^H$, whereby the restrictions hold whenever, for a given $\boldsymbol{\beta}$, the ICFs are nonnegative for all prices in the relevant region $\boldsymbol{\psi}$,

$$\mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \;\forall\; \mathbf{p} \in \boldsymbol{\psi}.$$

For example, if $f(\mathbf{p};\boldsymbol{\beta})$ is a twice continuously differentiable, linear homogenous in $\mathbf{p}$ unit cost function with $K = 3$ input prices, then the ICFs can be defined as[6]

$$i_1 = \partial f(\mathbf{p};\boldsymbol{\beta})/\partial p_1, \qquad i_2 = \partial f(\mathbf{p};\boldsymbol{\beta})/\partial p_2, \qquad i_3 = \partial f(\mathbf{p};\boldsymbol{\beta})/\partial p_3,$$

$$i_4 = -\,\partial^2 f(\mathbf{p};\boldsymbol{\beta})/\partial p_1^{\,2} \qquad \text{and} \qquad i_5 = \begin{vmatrix} \partial f^2(\mathbf{p};\boldsymbol{\beta})/\partial p_1 \partial p_1 & \partial f^2(\mathbf{p};\boldsymbol{\beta})/\partial p_1 \partial p_2 \\ \partial f^2(\mathbf{p};\boldsymbol{\beta})/\partial p_2 \partial p_1 & \partial f^2(\mathbf{p};\boldsymbol{\beta})/\partial p_2 \partial p_2 \end{vmatrix}.$$

Note that previous *global* and *local* estimation methodologies differ in the way $\boldsymbol{\psi}$ is defined. If $\mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \;\forall\; \mathbf{p} \in \boldsymbol{\psi}$, we say that regularity is imposed (i) *locally* if $\boldsymbol{\psi}$ consists of one or more singular disconnected points in $\boldsymbol{\pi}$, (ii) *globally* if $\boldsymbol{\psi} = \boldsymbol{\pi}$, and (iii) *regionally* if $\boldsymbol{\psi}$ is some connected subset of $\boldsymbol{\pi}$. Given the trade off between *flexibility*, on the one hand, and *regularity violations* on the other, we follow the idea of GALLANT and GOLUB (1984) and consider imposing the conditions *regionally*. For this purpose we now define a particularly relevant $\boldsymbol{\psi}$.

> ***Definition 1:*** *The empirically relevant subset of* $\boldsymbol{\pi}$, $\boldsymbol{\psi}$, *is a closed and connected*[7] *subset of* $\boldsymbol{\pi}$ *that covers the empirically relevant price space, defined as containing all sample observation* $n = 1,\ldots,N$ *as well as any price points* $c = 1,\ldots,C$ *that will be used for subsequent analyses and/or simulations based on the estimated model.*

---

[6] Note that nonnegativity of i_1, i_2 and i_3 imposes monotonicity. Nonnegativity of i_4 and i_5 imposes the curvature condition by signing the first and second principal minors of the $K \times K$ Hessian $\partial^2 f(\mathbf{p};\boldsymbol{\beta})/\partial\mathbf{p}\partial\mathbf{p}'$. Since by linear homogeneity of $f(\cdot)$ the Hessian has rank $K - 1$, it is not necessary to generate an additional ICF to account for the K^{th} principal minor.

[7] A connected set is such that any two points in the set can be connected by a path of straight segments totally contained in the set. Formally: let $\mathbf{S}$ be a topological space. $\mathbf{X} \subset \mathbf{S}$ is connected iff we cannot find open sets $\mathbf{U}$, $\mathbf{V} \subset \mathbf{X}$ such that $\mathbf{U} \cap \mathbf{V} = \varnothing$ and $\mathbf{U} \cup \mathbf{V} = \mathbf{X}$. The requirement that $\boldsymbol{\psi}$ is a closed set simplifies the proofs of some later propositions, but is not necessary for any other reason.

In contrast to previous practice, we here require $\boldsymbol{\psi}$ to be a connected set. It rules out the possibility that any small irregular region in between two disconnected regular regions can destroy overall regularity (see fig. 1).

-- INSERT FIG. 1 --

2.2. Statistical model and Bayesian context

Let

$$\mathbf{y} = \mathbf{f}(\mathbf{P};\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{1}$$

be the statistical model of interest, whereby $\mathbf{y}$ is an $M{\cdot}N \times 1$ vector of N observations on M endogenous variables, which represent transformations of $N \times K$ observed prices $\mathbf{P}$, and $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ is an $L \times 1$ unknown parameter vector.[8] We assume that $\boldsymbol{\varepsilon}$ is an $M{\cdot}N \times 1$ unknown error vector with mean $\mathrm{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Further, $\boldsymbol{\Theta}$ is the L-dimensional parameter space, which, if the regularity conditions are to hold for all values of $\mathbf{p}$ in $\boldsymbol{\psi}$, reduces to the L-dimensional regular subset $\boldsymbol{\Theta}^{\mathrm{R}} \subset \boldsymbol{\Theta}$ defined as[9]

$$\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi} = \{\boldsymbol{\beta}: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \;\forall\; \mathbf{p} \in \boldsymbol{\psi}\}. \tag{2}$$

The marginal posterior distribution for $\boldsymbol{\beta}$ is derived by applying Bayes rule

$$p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto \int \mathscr{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y}) \cdot p(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{\psi}) \mathrm{d}\boldsymbol{\Sigma} \tag{3}$$

where $\mathscr{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})$ is the likelihood function summarizing the sample information, $p(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{\psi})$ is the joint prior distribution on the parameters, given $\boldsymbol{\psi}$, and $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ is the conditional posterior. Assuming the standard ignorance prior on the covariance matrix, $p(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(M*+1)/2}$, and further assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are independent, the joint prior is defined as

$$p(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{\psi}) = p(\boldsymbol{\beta}|\boldsymbol{\psi}) \cdot |\boldsymbol{\Sigma}|^{-(M+1)/2}. \tag{4}$$

---

[8] Note that the matrix denoted by the capital letter $\mathbf{P}$ represents n observations on the lower case price vector $\mathbf{p} = [p_1, p_2, \ldots, p_K]^{\mathrm{T}}$.

[9] We use the superscript '$^{\mathrm{R}}$' for a 'regular' set, and '$^{\mathrm{IR}}$' for an 'irregular' set. E.g. for the irregular parameter space we write $\boldsymbol{\Theta}^{\mathrm{IR}}$. Note that generally for *any given* connected or disconnected set $\boldsymbol{\psi}_*$, $\boldsymbol{\Theta}$ consists of two disjoint subsets, such that $\boldsymbol{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}_* \cup \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_* = \boldsymbol{\Theta}$.

In the remainder of the paper we do not impose any additional information in our prior other than that needed to account for the economic theory constraints imposed on $\boldsymbol{\psi}$. Recognizing that the definition of the regular parameter set $\Theta^R|\boldsymbol{\psi}$ is dependent on the choice of $\boldsymbol{\psi}$, the marginal conditional improper[10] prior on the β vector is specified as an indicator function

$$p(\beta|\boldsymbol{\psi}) = I_{\Theta^R|\boldsymbol{\psi}}(\boldsymbol{\beta}) \tag{5}$$

where the prior equals 1 if regularity holds at the value $\beta \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$, and equals 0 otherwise.

The notation used in (1)-(5) highlights the conditionality upon $\boldsymbol{\psi}$ because it not only determines the applicable domain for $\mathbf{f}(\mathbf{p};\beta)$ but also determines the shape of $\Theta^R|\boldsymbol{\psi}$ and therefore the potential fit of the economic model to the data. In the remainder of the paper $p(\beta|\mathbf{y},\boldsymbol{\psi})$ denotes the *regularity posterior* containing all of the information about the parameters that can be extracted from a) economic theory, b) data and c) the chosen model, $\mathbf{y} = \mathbf{f}(\mathbf{P};\beta) + \varepsilon$, as applicable to a given empirically relevant region $\boldsymbol{\psi}$ of input price space.

2.3. Markov Chain Monte Carlo and Accept-Reject algorithm

We now turn towards the simulation technique used to generate outcomes from the regularity posterior $p(\beta|\mathbf{y},\boldsymbol{\psi})$, which are then used to obtain point estimates and to draw posterior inferences. One possible method is to approximate posterior expectations numerically by applying a Markov Chain Monte Carlo technique. For example, a Metropolis-Hastings algorithm can be used to generate J (pseudo-) random outcomes, $\mathbf{b}^{(j)}$, $j = 1,\ldots,J$ from $p(\beta|\mathbf{y},\boldsymbol{\psi})$ on the support Θ^R. The outcomes are then used to approximate posterior expectations via the appropriate empirical

[10] Note that typically a prior distribution is a function of the parameters only and has the entire parameter space as its domain. In our case however $p(\beta|\boldsymbol{\psi})$ also includes information about the price space as part of its specification. Also, $I_{\Theta^R|\boldsymbol{\psi}}(\boldsymbol{\beta})$ is technically not a "proper" prior distribution. It is not normalized to integrate to 1, and moreover, if $\Theta^R|\boldsymbol{\psi}$ does not have finite volume, $\int p(\beta|\boldsymbol{\psi})d\beta = \infty$. However our prior effectively indicates the set membership of β, i.e., if it is regular or not, and it is an uninformative prior on $\Theta^R|\boldsymbol{\psi}$.

estimates, e.g. $J^{-1}\sum_{j=1}^{J} g(\mathbf{b}(j))$ for approximating $E[g(\beta)]$. The estimates converge to the true expectations as J increases. [11]

To account for the regularity prior $p(\beta|\psi)$, the simulator should ensure that any drawn parameter vector $\mathbf{b}^{(j)}$ implies regularity of $\mathbf{f}(\mathbf{p};\beta)$ for every point $\mathbf{p}$ in the predefined set ψ, i.e. $\mathbf{b}^{(j)} \in \Theta^R|\psi \; \forall \; j$. Since there are an infinite number of points in ψ, they cannot all be checked explicitly. In general the connectedness can be approximated by a fine grid denoted by the disconnected set $\psi_g \subset \psi$ which consists of possibly tens-of-thousands of equidistant distinct points.[12] Within the MCMC an Accept-Reject Algorithm is then implemented to guarantee that $\forall \; \mathbf{b}^{(j)}$ the regularity conditions hold for any single grid point, i.e. that $\mathbf{b}^{(j)} \in \Theta^R|\psi_g \; \forall \; j$, whereby $\Theta^R|\psi_g$ is the *approximated* regularity posterior support, which will tend towards the actual set $\Theta^R|\psi$ the finer the approximation grid ψ_g. In order to circumvent the approximate nature of this representation, in a later subsection we identify problem conditions under which checking certain key points in ψ will guarantee overall regularity $\forall \; \mathbf{p} \in \psi$.

3. Regionally regular estimation procedure

This section describes our proposed method for estimating $f(\mathbf{p};\beta)$ subject to the nonlinear inequality constraints $\mathbf{i}(\mathbf{p};\beta) \geq \mathbf{0} \; \forall \; \mathbf{p} \in \psi$. To start we provide a complete stepwise description in box 1. The procedure consists of three parts: pre-analysis of the problem (step 1 to step 4), application of the MHARA (step 5 to step 11) and inferences based on the regularity posterior (step 12). In the subsections to follow, we explain the objectives of the steps that are nonstandard[13] and develop necessary technical details.

---

[11] See literature cited in footnote 13 for useful introductions into MCMC methods.

[12] I.e. in the case of a hyperrectangle ψ_g is defined as a) selecting Q equidistant values between the vertices of ψ, p_k^{min} and p_k^{max} as $p_k^q = p_k^{min} + (q\text{-}1)Q^{-1}(p_k^{max} - p_k^{min}) \; \forall \; q \in \{1,\ldots, Q\}$ and using all possible $Q \cdot K$ combinations of prices to generate ψ_g.

[13] ***Step 1, Step 5, Step 10*** and ***Step 11*** are not further elaborated on because their content is either obvious from the explanation given in box 1, or they are part of the conventional Metropolis-Hastings algorithm, which we assume the reader to be familiar with. In order to

Box 1: The 12-step procedure – pre-Analyses (1)-(4), MAHRA (5)-(11), inference (12)

Step 1	Estimate $\mathbf{y} = \mathbf{f}(\mathbf{P};\beta) + \varepsilon$ *without* imposing inequality constraints to obtain the unconstrained estimate $\mathbf{b}_u$ of β as well as the estimated $L \times L$ covariance matrix $\mathbf{cov}(\mathbf{b}_u)$.
Step 2	Define the ICFs that characterize the regularity conditions for the function being estimated.
Step 3	Define ψ according to definition 1. If the proposed region is not convex, define a sequence of I convex subsets ψ_i such that $\psi = \bigcup_{i=1}^{I} \psi_i$.
Step 4	Selection of evaluation points: Analyze for the h^{th} ICF, $i_h(\mathbf{p};\beta)$, which *properties* I to *property* V hold $\forall\ (\mathbf{p},\beta) \in (\psi \times \Theta)$ and define ψ_{gh} according to table 1. Repeat step 4 $\forall\ h$**.**
Step 5	Initialize the Markov Chain with a regular parameter vector: If $\mathbf{b}_u \in \Theta^R$, set $\mathbf{b}^{(0)} = \mathbf{b}_u$ else $\mathbf{b}^{(0)} = \mathbf{0}$. Set $j = 0$.
Step 6	Generate a candidate $\mathbf{b}^{(*)}$ by the proposal distribution $\delta \cdot p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$, whereby δ is to be set so that approximately 25%-50% of the regular draws $\mathbf{b}^{(*)}$ become accepted in step 10.
Step 7	If $\mathbf{b}^{(*)}$ is irregular at the vertices of ψ, go to step 6.
Step 8	Repeat step 4, but instead of evaluating the ICFs conditional on $(\mathbf{p},\beta) \in (\psi \times \Theta)$, evaluate the ICFs $\forall\ (\mathbf{p},\mathbf{b}^{(*)}) \in (\psi \times \mathbf{b}^{(*)})$, i.e. conditional on the last draw $\mathbf{b}^{(*)}$.
Step 9	If $\mathbf{b}^{(*)}$ is regular in ψ_g, calculate $r = p(\mathbf{b}^{(*)}\|\mathbf{y},\psi)/p(\mathbf{b}^{(j)}\|\mathbf{y},\psi)$, else go to step 6.
Step 10	if $r > 1$, $\mathbf{b}^{(j+1)} = \mathbf{b}^{(*)}$ else if Uniform(0,1) ≤ r, $\mathbf{b}^{(j+1)} = \mathbf{b}^{(*)}$, else $\mathbf{b}^{(j+1)} = \mathbf{b}^{(j)}$.
Step 11	Increment j by $j = j+1$. Go to step 6, until $j = J+S$, whereby $\{\mathbf{b}^{(j)}\}_{j=1}^{S}$ are the burn-in draws to be discarded after the final loop such that $\{\mathbf{b}^{(j)}\}_{j=S+1}^{J+S}$ are the outcomes to be considered for constructing $p(\beta\|\mathbf{y},\psi)$.
Step 12	Analyze $p(\beta\|\mathbf{y},\psi)$, i.e. calculate point estimates and perform inferences.

The dotted arrows indicate backward jumps in the algorithm which are conditional on the fact that the last drawn parameter vector $\mathbf{b}^{(*)}$ is irregular. The number of times these jumps occur is unknown prior to the estimation. In contrast, the loop indicated with the solid arrow is proceeded $J+S$ times.

3.1. Pre-Analysis: selection of regular region and approximation grid

The pre-analysis provides necessary information for the subsequent application of the MHARA especially the definition of the prior distribution $p(\beta,\psi) = I_{\Theta^R|\psi}(\beta)$: The regularity conditions (defined by economic theory) are identified (step 2), the empirical relevant region ψ is chosen by the researcher (step 3) and subsequently approximated by a grid ψ_g (step 4).

keep it is as uncomplicated as possible we outline the simplest way of implementing the Markov Chain. Other procedures like multiple chains and other proposal distributions are suggested in the literature. The reader is referred to CHIB and GREENBERG (1996), RICHARSON and SPIEGELHALTER (1996), ROBERT and CASELLA (1999) or CHEN, SHAO and IBRAHIM (2000) for a further discussion of appropriate modifications of the Metropolis-Hastings algorithm.

Step 2: The necessary regularity conditions of $f(\cdot)$ are to be translated into H ICFs, $\mathbf{i} \equiv [i_1, i_2 \ldots, i_H]$, such that economic theory holds whenever $\mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0}$. An illustrative example for the case of monotonicity and curvature restrictions was given in section 2.1.

Step 3: In contrast to defining $\boldsymbol{\psi}$ as one convex hyperrectangle (as in TERRELL 1996), it can be advantageous to define $\boldsymbol{\psi}$ as any connected (possibly non-convex) set. In order to see this, consider first the following adaptation of a well-known result from optimization theory:

> ***Lemma 1***:[14] *Let $\boldsymbol{\psi}_*$ be any subset of the regressor space $\boldsymbol{\pi}$ and let s: $\boldsymbol{\Theta}^R|\boldsymbol{\psi}_* \rightarrow \mathbb{R}^1$ be any scalar function.*
>
> $$\textit{If } \boldsymbol{\psi}_{1*} \subset \boldsymbol{\psi}_{2*}, \textit{ then } \max_{\boldsymbol{\beta} \in \boldsymbol{\Theta}^R | \boldsymbol{\psi}_{*1}} s(\boldsymbol{\beta}) \geq \max_{\boldsymbol{\beta} \in \boldsymbol{\Theta}^R | \boldsymbol{\psi}_{*2}} s(\boldsymbol{\beta}) .$$

Suppose $s(\boldsymbol{\beta})$ is any scalar goodness of fit measure maximized when estimating the model. The lemma then states that the resulting fit of the estimated model imposing regularity in $\boldsymbol{\psi}_{1*}$ is at least as good as the fit imposing regularity in $\boldsymbol{\psi}_{2*}$, given that $\boldsymbol{\psi}_{1*} \subset \boldsymbol{\psi}_{2*}$. This suggests defining $\boldsymbol{\psi}$ as small as possible. Instead of defining a hyperrectangle, $\boldsymbol{\psi}$ could be reduced to only cover N data and any additional C points used in subsequent analysis of the model. In order to apply the subsequent methodology, the resulting in general non-convex $\boldsymbol{\psi}$ needs to be decomposed into I *convex* subsets $\boldsymbol{\psi}_i \; \forall \; i = 1,\ldots,I$, such that $\boldsymbol{\psi} = \bigcup_{i=1}^{I} \boldsymbol{\psi}_i$.[15] In the context of applying the methodology (see section 4) it turns out that it is practical to construct $\boldsymbol{\psi}$ as $I = N+C$ line segments connecting all empirically relevant points thereby promising an increased fit of the estimated model to the data.

---

[14] The proof follows immediately from the definition of $\boldsymbol{\Theta}^R|\boldsymbol{\psi}_* = \{\boldsymbol{\beta}: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \; \forall \; \mathbf{p} \in \boldsymbol{\psi}_*, \boldsymbol{\beta} \in \boldsymbol{\Theta}\}$ which implies that ceteris paribus, the larger the constraining set $\boldsymbol{\psi}_* \subset \boldsymbol{\pi}$, the smaller is the support $\boldsymbol{\Theta}^R$, i.e. if $\boldsymbol{\psi}_{1*} \subset \boldsymbol{\psi}_{2*}$, then $\boldsymbol{\Theta}^R|\boldsymbol{\psi}_{1*} \supset \boldsymbol{\Theta}^R|\boldsymbol{\psi}_{2*}$. Consequently, due to smaller $\boldsymbol{\Theta}^R$, the less flexible is the function, i.e. the less is the ability of the function to fit the data.

[15] Since some nonconvex supersets cannot be decomposed into a finite union of convex subsets, the requirement to define each subset $\boldsymbol{\psi}_i$ to be convexly shaped limits the generality of the construction of possible regular regions. However, such nonconvex sets can be arbitrarily well approximated for large I. For applied work we propose nonconvex sets which circumvent this problem, see the "string approach" in section 4.2.

Whereas step 3 focused on the selection of $\boldsymbol{\psi}$, the next issue concerns the construction of the evaluation grid $\boldsymbol{\psi}_g$, which is conditional on a given set $\boldsymbol{\psi}$.

Step 4: As outlined in section 2.3, $\boldsymbol{\psi}$ is approximated by $\boldsymbol{\psi}_g$ and regularity is explicitly checked for a high number, say Q, of grid points. It remains uncertain, however, if the selected *Q-grid* is dense enough to avoid irregularity that may occur in between grid points.

The purpose of step 4 is to identify conditions under which it will be guaranteed that if certain key areas or singular points in $\boldsymbol{\psi}$ are regular, then other areas of interest are regular as well. This may allow for a reduction of regularity checks to a number $Q^* < Q$ that

a) improve the computational speed of the algorithm and

b) maintain the accuracy of the approximation obtained from the original Q-grid.

In order to identify those conditions the following properties relating to $f(\mathbf{p};\beta)$, $\boldsymbol{\psi}$, and i_h are exploited:

Property I: i_h *has property* I, *iff each of the K derivatives*, $\partial i_h/\partial p_k$, *is continuous and either* $\leq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$ *or* $\geq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$. *The signs may however be different across the K derivatives*.

Property II: $\boldsymbol{\psi}$ *is a closed and connected hyperrectangle constructed such that each of its sides is parallel to one of the K price-axes*.

Property III: i_h *has property* III, *iff the derivative with respect to at least one price (say the* m^{th} *price) is continuous and is either* $\partial i_h/\partial p_m \geq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$ *or* $\partial i_h/\partial p_m \leq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$.

Property IV: i_h *is quasiconcave in* $\mathbf{p}$ *and* $\boldsymbol{\psi}$ *is convex.*

Property V: $f(\mathbf{p};\beta)$ *is twice continuously differentiable and homogenous in* $\mathbf{p}$.

Table 1 below summarizes six cases for constructing sufficient "evaluation sets", $\boldsymbol{\psi}_h$, for the h^{th} ICF, where the $\boldsymbol{\psi}_h$'s are proper subsets of $\boldsymbol{\psi}$. Depending on the *properties* I-V $\boldsymbol{\psi}_h$ can take 5 different forms defined as follows:

(1) $\mathbf{B}_h = \text{bd}(\boldsymbol{\psi})$ denotes the boundary of $\boldsymbol{\psi}$.

(2) The $K \times 1$ price vector $\mathbf{z}_h$ is one vertex of the hyperrectangle $\boldsymbol{\psi}$. Given the proof of proposition 1b in the appendix, which vertex out of the 2^K vertices must be explicitly

checked (for the sign of i_h) depends on the signs of the derivatives of the ICF: If $\partial i_h/\partial p_k \leq 0$ $\forall$ $\mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\max}$ and if $\partial i_h/\partial p_k \geq 0$ $\forall$ $\mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\min}$.

(3) $\mathbf{Z}_h = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{2^K}]_h$ is a $K \times 2^K$ matrix of all vertices of the hyperrectangle $\boldsymbol{\psi}$.

(4) $\mathbf{S}_h \subset \mathbf{B}$ is one side of the hyperrectangle. Considering the proof of proposition 1b and corollary 2b in the appendix it follows that $\mathbf{S}_h$ is orthogonal to the m^{th} price-axis. Further details on the construction of the grid $\mathbf{S}_{gh}$ are given in the appendix.

(5) $\mathbf{S}^* \subset \mathbf{B}$ is the set, which can be described as a "shield" masking $\boldsymbol{\psi}$ from below (i.e. the origin $\mathbf{0} \in \boldsymbol{\pi}$, see the illustrations in Fig. 2). In order to define $\mathbf{S}^*$, let $l(\mathbf{0},\mathbf{y})$ be a straight line through the origin $\mathbf{0}$ and through $\mathbf{y} \in \boldsymbol{\pi}$, then $\mathbf{S}^* = \{\mathbf{p} \in \text{bd}(\boldsymbol{\psi})$: $\forall$ φ if $\varphi \in \text{bd}(\boldsymbol{\psi}) \cap l(\mathbf{0},\mathbf{p})$, then $\|\mathbf{p}\| \leq \|\varphi\|\}$.

-- INSERT FIG. 2 ---

- INSERT TABLE 1 --

Considering the above five definitions of the possible forms of $\boldsymbol{\psi}_h$, the six cases in table 1 should be read row-wise as follows:

For cases 1 – 5: *Suppose for the h^{th} elementary ICF, i_h, the properties (designated by +) hold:* $i_h \geq 0$ $\forall$ $\mathbf{p} \in \boldsymbol{\psi}$ *iff* $i_h \geq 0$ $\forall$ $\mathbf{p} \in \boldsymbol{\psi}_h$ *(whereby $\boldsymbol{\psi}_h$ takes the form as indicated in the column '$\boldsymbol{\psi}_h$').*

Case 6: *Suppose property V holds. Then for all ICFs* $\mathbf{i}^*(\cdot)$ *that impose nonnegative slope, nonpositive slope, concavity and/or convexity:* $\mathbf{i}^*(\cdot) \geq \mathbf{0}$ $\forall$ $\mathbf{p} \in \boldsymbol{\psi}$ *iff* $\mathbf{i}^*(\cdot) \geq \mathbf{0}$ $\forall$ $\mathbf{p} \in \mathbf{S}^*$.

The first five cases are independent of the type of regularity conditions to be imposed. Case 6 is less general but applies to all ICFs which impose monotonocity and curvature, (and thus suits the cost-function example) in which case only the shield $\mathbf{S}^*$ has to be evaluated.

Of particular interest are the cases 2 and 5, which hold frequently when enforcing the monotoniciy and first principal minor constraints. In these cases $\boldsymbol{\psi}_h$ is defined as one vertex $\mathbf{z}_h$

(case 2) or all vertices $\mathbf{Z}_h$ (case 5) of $\boldsymbol{\psi}$, which leads to the maximum reduction in the number of explicit regularity checks, enhancing the computational speed of MHARA substantially. For example in case 5, if $\boldsymbol{\psi}$ is a hyperrectangle and $K = 3$, then $Q^* = 2^K = 8 < Q$.

In practice all infinite $\boldsymbol{\psi}_h$ must be approximated by an h^{th} evaluation grid $\boldsymbol{\psi}_{gh}$. For example, the boundary evaluation set $\mathbf{B}_h = \text{bd}(\boldsymbol{\psi})$ is approximated by an evaluation grid $\mathbf{B}_{gh} \subset \mathbf{B}$, and $\mathbf{S}_h$ and $\mathbf{S}_h^*$ are approximated by $\mathbf{S}_{gh}$ and $\mathbf{S}_{gh}^*$ respectively. Conversely $\mathbf{z}_h$ and $\mathbf{Z}_h$ are *finite* evaluation sets that do not require the approximation subindex '$_g$'. This leads to the following useful result:

Proposition 5:[16] *If for all* $\mathbf{b}^{(j)}$ *case 2 or 5 hold* $\forall$ *h, then* $\forall\ \mathbf{p} \in \boldsymbol{\psi}$ $f(\mathbf{p};\mathbf{b}^{(j)})$ *is regular.*

3.2. The Metropolis-Hastings Accept-Reject algorithm and mitigating posterior bias

Steps 6 to 11 of the procedure apply the MHARA, which provides J random draws from the regularity posterior $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$. We elaborate on some of these steps below.

Step 6: $\mathbf{b}^{(*)}$, a candidate for the j^{th}+1 vector in the MCMC sequence $\{\mathbf{b}^{(j)}\}_{j=1}^{J+S}$, is generated by a symmetric *proposal distribution* $\delta \cdot p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$.[17] One possibility for drawing outcomes from $p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$ that accounts for *linear equality* constraints on parameters (e.g. for the symmetry condition on the Hessian $\partial^2 f(\mathbf{p};\boldsymbol{\beta})/\partial\mathbf{p}\partial\mathbf{p}'$) is to use the multivariate normal distribution $\mathbf{N}(\mathbf{b}^{(j)},\mathbf{cov}(\mathbf{b}_u))$ to generate the $L \times 1$ vector $\mathbf{b}^{(**)}$, and then to calculate

$$\mathbf{b}^{(*)} = \mathbf{b}^{(**)} - \mathbf{cov}(\mathbf{b}_u)\cdot\mathbf{R}^{\mathsf{T}}\cdot(\mathbf{R}\cdot\mathbf{cov}(\mathbf{b}_u)\cdot\mathbf{R}^{\mathsf{T}})^{-1}\cdot(\mathbf{R}\cdot\mathbf{b}^{(**)} - \mathbf{r}),$$

whereby $\mathbf{R}$ is a $V \times L$ design matrix and $\mathbf{r}$ is a $V \times 1$ vector chosen appropriately to impose V linear equality restrictions on $\mathbf{b}^{(*)}$.

---

[16] The proof follows directly from the propositions 1b and proposition 3 and noting that if the evaluation sets are finite, the regularity posterior can be simulated with support $\Theta^R|\psi = \Theta^R|\psi_g$, i.e., regularity is guaranteed on the connected set $\forall\ \mathbf{p} \in \boldsymbol{\psi}$ and there is no reliance on an arbitrary approximation grid.

[17] The term *proposal distribution* stems from the fact that $\delta \cdot p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$ *proposes* a new candidate $\mathbf{b}^{(*)}$ for the next state $\mathbf{b}^{(j+1)}$. Generally the proposal distribution is defined to be symmetric around the previous accepted point $\mathbf{b}^{(j)}$, in which case the tuning parameter δ is to be set that between 25%-50% of the regular draws $\mathbf{b}^{(*)}$ are accepted in step 10. The optimal acceptance rate depends on the number of parameters estimated, see ROBERT and CASELLA (pp.281-283: 2002) for a recent discussion.

Step 7 and 9: Step 7 is inserted to save computing time associated with step 8 for vectors $\mathbf{b}^{(*)}$ that are already irregular at the vertices of $\boldsymbol{\psi}$. If $\mathbf{b}^{(*)}$ is identified to be irregular (either after step 7 or 9), $\mathbf{b}^{(*)}$ must be *discarded* and a new $\mathbf{b}^{(*)}$ drawn in step 6 (see the dotted arrows in box 1) using the last regular draw $\mathbf{b}^{(j)}$ as the mean of the symmetric proposal distribution $\delta \cdot p(\mathbf{b}^{(*)},\mathbf{b}^{(j)})$. This is repeated until $\mathbf{b}^{(*)} \in \Theta^{R}|\boldsymbol{\psi}_{g}$. The 'discarding' is necessary to avoid an upward bias of the regularity posterior density values in the neighborhood of the truncation boundary.[18]

To our knowledge in all previously published descriptions of the MHARA[19] it was common to *repeatedly* include the last regular $\mathbf{b}^{(j)}$ as an outcome of the simulated regularity posterior as $\mathbf{b}^{(j+1)} = \mathbf{b}^{(j)}$ until $\mathbf{b}^{(*)} \in \Theta^{R}|\boldsymbol{\psi}$. This practice, however, distorts the simulated regularity posterior in the peripheral region of $\Theta^{R}|\boldsymbol{\psi}$ close to the truncation boundary to $\Theta^{IR}|\boldsymbol{\psi}$. This is due to the fact that the probability of drawing an irregular $\mathbf{b}^{(*)}$ is higher, the closer the last regular draw $\mathbf{b}^{(j)}$ is to the frontier of $\Theta^{IR}|\boldsymbol{\psi}$.[20]

To complete step 9, if the drawn parameter vector $\mathbf{b}^{(*)}$ is regular $\forall\ \mathbf{p} \in \boldsymbol{\psi}_{g}$, calculate[21]

$$r = p(\mathbf{b}^{(*)}|\mathbf{y},\boldsymbol{\psi})/p(\mathbf{b}^{(j)}|\mathbf{y},\boldsymbol{\psi}). \tag{6}$$

Finally note that step 7 and the 'else condition' of step 9 (see the dotted arrows in box 1) *approximate* the behavior of the indicator function $I_{\boldsymbol{\Theta}^{R}|\boldsymbol{\psi}}(\boldsymbol{\beta})$ by subtracting $\Theta^{IR}|\boldsymbol{\psi}_{g}$ (instead of $\Theta^{IR}|\boldsymbol{\psi}$) from Θ. In order for MHARA to more closely emulate $I_{\boldsymbol{\Theta}^{R}|\boldsymbol{\psi}}(\boldsymbol{\beta})$, step 8 can be implemented.

---

[18] Since the bias arises independently if sampling from $\Theta^{R}|\boldsymbol{\psi}$ or from $\Theta^{R}|\boldsymbol{\psi}_{g}$, we will drop the subindex 'g' for the explanation.

[19] Among others, the studies of O'DONNELL, SHUMWAY and BALL (1999), GRIFFITHS, O'DONNELL and TAN CRUZ (2000), GRIFFITHS (2001), CHUA, GRIFFITHS and O'DONNELL (2001), and CUESTA ET AL. (2001), O'DONNELL, RAMBALDI and DORAN (2001) did not account for this bias.

[20] Denote the relevant peripheral region close to or on the boundary $\Theta^{IR}|\boldsymbol{\psi}$ as $\Theta^{b}|\boldsymbol{\psi}$ and denote the simulated posterior as *p_hat*. Then the bias arises of the form $p\_hat(\boldsymbol{\beta}^{b}|\mathbf{y},\boldsymbol{\psi}$, *without 'discarding'*$) > p(\boldsymbol{\beta}^{b}|\mathbf{y},\boldsymbol{\psi})$ for $\boldsymbol{\beta}^{b} \in \Theta^{b}|\boldsymbol{\psi}$. A numerical example illustrating the bias by comparing the previous to the above simulation technique can be found in WOLFF et al. (2003).

[21] E.g. in the case of a normal SUR model (6) becomes $[|(N\text{-}L)\boldsymbol{\Sigma}^{(*)}|/|(N\text{-}L)\boldsymbol{\Sigma}^{(j)}|]^{-N/2}$ which can be derived from the definition of the unconstrained posterior $p_{u}(\boldsymbol{\beta}|\mathbf{y}) \propto \int\!\!\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})|\boldsymbol{\Sigma}|^{-(M+1)/2}\mathrm{d}\boldsymbol{\Sigma}$ and the fact that it is directly proportional to $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ by $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto |(N\text{-}L)\boldsymbol{\Sigma}|^{N/2}$, (ZELLNER, 1971:p.243). Cancelling out the normalizing constants and factoring out the exponents $^{-N/2}$ yields $[|(N\text{-}L)\boldsymbol{\Sigma}^{(*)}|/|(N\text{-}L)\boldsymbol{\Sigma}^{(j)}|]^{-N/2}$.

Step 8: The same procedure applies as in step 4, with the modification that $f(\cdot)$ and $\mathbf{i}(\cdot)$ are evaluated conditionally on the drawn parameter vector $\mathbf{b}^{(*)}$. To save computing time, if $\boldsymbol{\psi}_{gh} = \mathbf{Z}_h$ or $\boldsymbol{\psi}_{gh} = \mathbf{z}_h$ in step 4, the h^{th} evaluation of step 8 can, of course, be skipped.

3.3. Point estimates: inconsistency of the mean and two alternatives

Step 12: Steps 1 to 11 generated J outcomes of $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}_g)$, which can now be used to derive point estimates and to draw posterior inferences. Finite sample inferences such as posterior moments and highest posterior density regions can be directly computed using well-known Monte Carlo techniques.

As far as we are aware, all previous studies applying MCMC and Importance sampling to impose regularity conditions define the point estimate of $\boldsymbol{\beta}$ as the mean $\mathrm{E}[\boldsymbol{\beta}]$ of the regularity posterior.[22] However, this may result in regularity violations, as indicated in the following proposition.

Proposition 6:[23] *Let* $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ *be the regularity posterior with parameter support* $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$. *If an inequality constraint is a nonlinear function of* $\boldsymbol{\beta}$, *then* $\mathrm{E}[\boldsymbol{\beta}] = \int \boldsymbol{\beta} \cdot p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \mathrm{d}\boldsymbol{\beta}$ *can reside in either* $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ *or* $\boldsymbol{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}$, *and thus* $f(\mathbf{p};\mathrm{E}[\boldsymbol{\beta}])$ *can lose the property of being regular for some* $\mathbf{p} \in \boldsymbol{\psi}$.

We propose two alternative estimators that, in addition to imposing regularity (objective I), maximize a model fit measure $s(\boldsymbol{\beta})$ on $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_g$, as indicated by Lemma 1 (objective II). Our first suggestion for an estimator is best motivated under the assumption of Gaussian noise. The second is motivated independently of the noise probability distribution.

Under the assumption of a normal error distribution, we suggest selecting the mode

---

[22] These include BARNETT, GEWEKE and WOLFE (1991), KOOP, OSIEWALSKI and STEEL (1994, 1997), TERRELL (1996), TERRELL and DASHTI (1997), O'DONNELL, SHUMWAY and BALL (1999), GRIFFITHS, O'DONNELL and TAN CRUZ (2000), CHUA, GRIFFITHS and O'DONNELL (2001), KLEIT and TERRELL (2001), CUESTA ET AL. (2001), ADKINS, RICKMAN and HAMEED (2002), O'DONNELL, RAMBALDI and DORAN (2001) and O'DONNELL and COELLI (2003).

[23] The proof follows directly by noting that for nonlinear inequality constraints the constraint set $\boldsymbol{\Theta}^{\mathrm{R}}$ is not necessarily convex. Hence linear combinations over $\boldsymbol{\Theta}^{\mathrm{R}}$ can reside outside of $\boldsymbol{\Theta}^{\mathrm{R}}$.

$$\beta^{(\text{mode})} = \underset{\beta \in \Theta^R | \psi_g}{\arg\max} \left\{ p(\beta \mid \mathbf{y}, \psi_g) \right\}.$$

of the regularity posterior as the point estimate to maximize model fit subject to the regularity conditions. To motivate $\beta^{(\text{mode})}$, note that the information contained in the normal unrestricted posterior $p_u(\beta|\mathbf{y}) \propto |(N\text{-}L)\Sigma|^{-N/2}$ (see ZELLNER 1971:p.243) is strictly monotonically related to the *generalized variance of the fit* $|\boldsymbol{\Sigma}|^{-1}$, which can be used as a goodness of fit indicator. In fact, BARNETT (1976) proved that the minimization of $|\boldsymbol{\Sigma}|$ is equivalent to Maximum Likelihood (ML) estimation in the case of the nonlinear normal classical SUR model. Since $(N\text{-}L)$ and the exponent $-N/2$ are fixed constants, the minimization of $|\boldsymbol{\Sigma}|$ over $\beta \in \Theta$ produces the exact same result as the maximization of $p_u(\beta|\mathbf{y})$ over $\beta \in \Theta$. So long as no other prior than the regularity prior is applied, we have that $p(\beta|\mathbf{y},\psi) \propto p_u(\beta|\mathbf{y}) \cdot I_{\Theta^R|\psi}(\beta) \propto |(N\text{-}L)\Sigma|^{-N/2}$ for $\beta \in \Theta^R|\psi$. Thus the normal classical inequality-constrained-ML estimator generates a point estimate that is numerically equivalent to the mode of $p(\beta|\mathbf{y},\psi)$. In order to approximate the solution based upon the MCMC outcomes $\{\mathbf{b}^{(j)}\}_{j=1}^{J+S}$, one can simply compare the values $p_u(\mathbf{b}^{(j)}|\mathbf{y})\ \forall\ j$ resulting from the MHARA as

$$\mathbf{b}^{(\text{mode})} = \underset{\mathbf{b}^{(j)}}{\text{argmax}} \left\{ |\ (N-L)\Sigma\ |^{-N/2(j)} \right\}.$$

An alternative estimator, which is not tied to Gaussian errors, can be based on a loss function (LF) criteria over $\Theta^R|\psi_g$. The estimator would be defined by solving

$$\beta^{(\text{LF}_\varphi)} = \underset{\beta^* \in \Theta^R|\psi_g}{\arg\min} \left\{ \int_{\beta \in \Theta^R|\psi_g} \| \beta^* - \beta \|_\varphi \ p(\beta \mid \mathbf{y}, \psi_g) d\beta \right\}$$

which minimizes the posterior weighted deviation over $\beta \in \Theta^R$, where $\|\cdot\|_\varphi$ is some vector norm[24] measuring the distance between two points within Θ^R. For example, with the standard Euclidean

[24] Given an N-dimensional $\mathbf{x}$ a general vector norm $\|\mathbf{x}\|_\varphi$, for $\varphi = 1,2,\ldots$ is a nonnegative defined as $\|\mathbf{x}\|_\varphi = [\Sigma_{n=1}^N |\mathbf{x}|^\varphi]^{1/\varphi}$. The special case $\|\mathbf{x}\|_\infty$ is defined as $\|\mathbf{x}\|_\infty = \max|x_n|$. The most commonly encountered vector norm is the Euclidian norm, given by $\|\mathbf{x}\|_2 = [\Sigma_{n=1}^N \mathbf{x}^2]^{1/2}$.

norm $\|\cdot\|_2$ $\mathbf{b}^{(\mathrm{LF}_2)} = \arg\min_{\mathbf{b}^{(j)}} \sum_{i=1}^{J} (\mathbf{b}^{(j)} - \mathbf{b}^{(i)})'(\mathbf{b}^{(j)} - \mathbf{b}^{(i)})$ which minimizes the empirical-MCMC analogue to the expected squared LF subject to the regularity constraints.

We reemphasize that if cases 2 or 5 of table 1 apply $\forall$ *h*, then $\mathbf{b}^{(\mathrm{LF}_\varphi)}$ and $\mathbf{b}^{(\mathrm{mode})}$ are members of the regular set $\Theta^{\mathrm{R}}|\boldsymbol{\psi}$ and hence both estimators are regularity-preserving (proposition 5). Conversely, if cases 2 and 5 do not hold, then without further knowledge one cannot exclude that the estimates belong to the irregular set $\Theta^{\mathrm{IR}}|\boldsymbol{\psi}$. Nevertheless, the following fact can be supportive: If $Q \to \infty$, i.e. the number of equidistant grid points of $\boldsymbol{\psi}_{\mathrm{g}}$ goes to infinity, and $\mathbf{i}(\cdot)$ is continuously differentiable, then any point estimate $\hat{\boldsymbol{\beta}} \in \Theta^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}$ is such that $f(\mathbf{p}; \hat{\boldsymbol{\beta}})$ is almost everywhere in $\boldsymbol{\psi}$ regularity-retaining.

The proposed methodology is general enough to be adopted in both the Bayesian and the Classical framework. In the Classical framework one could maximize a likelihood function subject to (non-)linear inequality constraints represented by the ICFs and the point estimate is the mode of the MCMC-simulated likelihood, which generally will be identical to the mode, $\boldsymbol{\beta}^{(\mathrm{mode})}$, of the regularity posterior. The suggested LF criterion, leading to $\boldsymbol{\beta}^{(\mathrm{LF}_\varphi)}$, is typically motivated from the Bayesian perspective and has no direct Classical analogue.

4. Numerical Examples

This section illustrates the proposed methodology by estimating a cost function subject to regularity conditions. For comparison purposes we re-estimate and extend some of the simulation experiments provided in the work of TERRELL (1995).[25] In the first subsection local, global and regional regularity approaches are compared based on a specified convex set $\boldsymbol{\psi}^{\square}$. The purpose of the second subsection is to demonstrate the effects of shrinking the size of $\boldsymbol{\psi}$.

[25] At this point we do not provide a real world application. Instead, the model is kept rather simple which simplifies notation and interpretation of the results related the imposition of the regularity conditions. However, generalizations are straightforward. E.g. output, as another explanatory variable, could be added while simultaneously imposing that *f* is convex and monotone increasing in output, as it is required by economic theory, in addition to the restrictions which are imposed with respect to **p**.

4.1. Experiment I - convex cube ψ

4.1.1. Data Generation

We now briefly describe the design of the simulations.[26] The *true* data generation process is formulated by the well-known CES cost function $f^{CES}(\mathbf{p};\alpha_k,\rho) = [\Sigma_{k=1}^{3} a_k^{1/(1-\rho)} \cdot p_k^{-\rho/(1-\rho)}]^{(1-\rho)/-\rho}$. As in TERRELL, no stochastic error term is added. The derivatives result, by Shephard's Lemma, in $K = 3$ input demand functions,

$$x_k = \partial f^{CES}/\partial p_k = [\alpha_k \cdot f^{CES}/p_k]^{1/(1-\rho)} \tag{7}$$

Following TERRELL, the data set for the first experiment (table 2) contains $N = 64$ observations, consisting of all combinations of the values 0.5, 0.8333, 1.1666 and 1.5 generated by $K = 3$ input prices. By (7) this produces $64 \cdot 3$ true input demand levels, where $\mathbf{x}_k$ is 64×1 with $k = 1, 2, 3$.

4.1.2. Estimation and Evaluation

The purpose of the first experiment is to assess potential advantages of the regional approach compared to the local and global approach both in terms of model fit and the propensity for regularity violations. The normal SUR system of $K = 3$ input demand functions, $\hat{\mathbf{x}}_k = \partial \mathbf{f}_k^{\mathbf{AIM}\tau)}(\mathbf{P};\hat{\boldsymbol{\beta}})/\partial p_k + \hat{\mathbf{u}}_k$ is estimated, whereby $\hat{\mathbf{u}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ represents the 64×1 approximation error vector to the 'true' data generation process (7), $L < N$ [27] and $\hat{\mathbf{x}}_k$ is the estimated k^{th} 64×1 input demand vector derived from the Asymptotically Ideal Production Model, AIM(τ), with

$$f^{AIM(1)} = \Sigma_{k=1}^{3} \beta_k p_k + \beta_4 p_1^{1/2} p_2^{1/2} + \beta_5 p_1^{1/2} p_3^{1/2} \beta_6 + \beta_6 p_2^{1/2} p_3^{1/2}$$

[26] For further details about the simulation set-up, the reader is referred to TERRELL (1995).

[27] This requirement is due to an important recent proof by GRIFFITHS, SKEELS and CHOTIKAPANICH (2002), ensuring a bounded solution for the unconstrained maximum likelihood function. They remark that heretofore most authors incorrectly assumed that $N > M$ and $N \geq \max\{L_m\}$ is sufficient, with L_m being the number of parameters of the m^{th} equation, $m = 1,\ldots,M$.

$$f^{\mathrm{AIM}(2)} = \Sigma_{k=1}^{3} \beta_k p_k + \beta_4 p_1^{3/4} p_2^{1/4} + \beta_5 p_1^{3/4} p_3^{1/4} + \beta_6 p_1^{1/2} p_2^{1/2} + \beta_7 p_1^{1/2} p_2^{1/4} p_3^{1/4} + \beta_8 p_1^{1/2} p_3^{1/2}$$
$$+ \beta_9 p_1^{1/4} p_2^{3/4} + \beta_{10} p_1^{1/4} p_2^{1/2} p_3^{1/4} + \beta_{11} p_1^{1/4} p_2^{1/4} p_3^{1/2} + \beta_{12} p_1^{1/4} p_3^{3/4} + \beta_{13} p_2^{3/4} p_3^{1/4}$$
$$+ \beta_{14} p_2^{1/2} p_3^{1/2} + \beta_{15} p_2^{1/4} p_3^{3/4},$$

which are homogenous of degree one, constant returns to scale unit cost functions.[28]

As in TERRELL (1995), the performance of the AIM(τ) is evaluated over the cubic region $\boldsymbol{\psi}^{\square} = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\}$ by defining a grid $\boldsymbol{\psi}^{\square}_{g} \subset \boldsymbol{\psi}^{\square}$ of 20 equidistant prices for each input. Thus in total $\boldsymbol{\psi}^{\square}_{g}$ consists of $Q = 20 \cdot 20 \cdot 20 = 8000$ points, $q = 1,\ldots,Q$. This grid is used to compute (a) the *maximum approximation error*, MAE = $\max\{\hat{u}_{gk}\}$, and (b) the *average absolute approximation error*, AAAE = $Q^{-1} \sum_{q=1}^{Q} |\hat{u}_{qk}|$, over all Q points, where $\hat{u}_{qk} = \hat{x}_{qk} - x_{qk}$ is the difference between the predicted input demand, estimated by the AIM(τ), and the (true) CES input demand of equation (7). Then pursuing our objective II of maximizing the model fit function $s(\beta)$ is equivalent to minimizing MAE or AAAE.

4.1.3. Results

-- INSERT TABLE 2 --

-- INSERT FIG. 3 --

The model fit measures, as well as the percentages of regularity violations of the grid points for the local, global and regional approach are displayed in table 2. In the first two columns we repeat TERRELL's (table 1 and 2, pp.9-10:1995) simulation experiment, and the last two columns apply the method described in section 4.

[28] A functional form is second order flexible, if it is capable of being *locally* equivalent to the true function in level, gradient, and Hessian at one given point in the price domain $\boldsymbol{\pi}$. This is the case for the AIM(1), which is equivalent to the well known Generalized Leontief. Through series expansions the order of flexibility can be increased to *locally* coincide with the true function at higher than second order derivatives. The AIM(2) maintains the flexibility order three. Asymptotically, $\tau \to \infty$, these forms converge *globally* to the true function. For a further discussion and definitions about second order flexibility see e.g. BARNETT (1983). For the concept and applications of globally flexible functional forms, see e.g. GALLANT and GOLUB (1984), TERRELL (1995) or BARNETT, GEWEKE and WOLFE (1991).

First the demand system is estimated subject to local concavity and monotonicity constraints guaranteeing regularity for the underlying AIM(τ) cost function at $\mathbf{p}^M = [1,1,1]$, i.e. at the mean of $\boldsymbol{\psi}^{\square}$. Compared to the other columns, the local approach provides the best model fit statistics but violates the regularity conditions in the neighbourhood of $\mathbf{p}^M$ (leading to regularity violations within $\boldsymbol{\psi}^{\square}$), which is illustrated in fig. 3. The (surprising) information is that the monotonicity violations are dominating the concavity violations. Given that TERRELL - and in fact most researchers in similar previous studies - did not check for monotonicity violations, this result should serve as a warning because any inferences based on such a cost function are suspect.

In the column 'global regularity' economic theory holds globally on π by, following TERRELL (1995), nonnegativity constraints on all the AIM parameters β, which confirms numerically the result of lemma 1 by showing a decreased model fit.

The last two columns show the MHARA[29] results imposing the regularity conditions regionally on $\boldsymbol{\psi}^{\square}$. First we take the mean – as is commonly done – as the point estimate for β. As one might expect this 'regional mean approach' leads to improved model fit measures compared to the global approach (e.g. a reduction of the AAAE by 33.6% and 69.7% and a reduction of the MAE by 38.1% and 71.6% in the case of the AIM(1) and AIM(2) respectively). However, only the mode, as the point estimate for β, guarantees regional regularity within $\boldsymbol{\psi}^{\square}$ (proposition 6). Results from the 'mode approach' are displayed in the last column of the table, confirming the theory outlined in section 3 that the model fit statistics are *always* superior to the 'mean approach', leading to a further reduction in the AAAE of 1.7% and 7.2% and to a reduction in the MAE of 8.7% and 2.3% for the AIM(1) and AIM(2), respectively.

Concerning the computational efficiency of the algorithm, it is worthwhile to note that instead of the full evaluation grid of 8000 points, due to the properties I to V, for all *H* ICFs the maximum of 1142 grid points of the set $\mathbf{S}_g^* \subset \boldsymbol{\psi}^{\square}{}_g$ had to be evaluated only. Furthermore, for the

[29] For MCMC sampling in the context of the normal SUR model, we want to refer to the very useful exposition by GRIFFITHS (2003).

AIM(1) often only one vertex had to be assessed. This significantly decreased the computational burden compared to previous approaches.

Summarizing table 2, imposing local regularity increases the model fit in all specifications at the cost of violating monotonicity and concavity within $\boldsymbol{\psi}$, which produces estimation results that are problematic in terms of economic interpretation and further analysis. Imposing regional regularity solves this problem and still significantly increases the model fit compared to the global approach. Moreover, apart from its appealing regularity preserving property, it seems relevant for model fit to use the mode instead of the mean.

4.2. Experiment II – comparison between convex and nonconvex $\boldsymbol{\psi}$

The purpose of this subsection is to analyze model performance for different definitions of $\boldsymbol{\psi}$ based on empirically relevant price sets.

-- INSERT TABLE 3 –

-- INSERT FIG. 4 --

The experimental design is based on the same (true) data generation process as in the previous subsection. However, instead of using the 64 observations, $N = 26$ data points are (randomly) selected from $\boldsymbol{\psi}^{\square} = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\}$, under the restriction that a) the smallest and the largest values are (again) elements of the boundary of $\boldsymbol{\psi}^{\square}$, i.e. $p_k^{\min} = 0.5 \; \forall \; k$ and $p_k^{\max} = 1.5 \; \forall \; k$ and that b) the points do not belong to three convex subsets that are eliminated from $\boldsymbol{\psi}^{\square}$. Suppose further that the purpose of the estimated model is to analyze $C = 4$ (policy) scenarios, and that the scenario prices are exogenously determined at 2 points within $\boldsymbol{\psi}^{\square}$ and at 2 points outside of $\boldsymbol{\psi}^{\square}$.[30] Then, a natural goal is to estimate the function such that all $N + C$ price points are regular (objective I) and that the model fit is as good as possible (objective II).

To evaluate the influence of different definitions of $\boldsymbol{\psi}$ the empirically relevant regions are chosen to be

(a) $\boldsymbol{\psi}^{\square}$, as before approximated by 8000 grid points $\boldsymbol{\psi}^{\square}_{g}$ and

(b) $\boldsymbol{\psi}^{string} = \bigcup_{i=1}^{29} \boldsymbol{\psi}_i$, which covers all 30 = I – 1 price points by connecting 29 straight lines $\boldsymbol{\psi}_i$, I = 1,...I, between $\mathbf{p}^M$ (which is one of the C selected scenario points) and each of the remaining $N + C - 1$ prices. We chose to approximate each line $\boldsymbol{\psi}_i$ by $\boldsymbol{\psi}_{ig}$ by taking 20 equidistant grid points between $\mathbf{p}^M$ and the i^{th} price point, leading to a total of 580 grid points for $\boldsymbol{\psi}_g$ only. Further, due to exploiting properties I-V, the evaluation grid could be reduced to 520 points, which is displayed in fig. 4. Furthermore, for the AIM(1), the grid could be further reduced to just 30 evaluation points, $\mathbf{Z}_h$, for assessing monotonicity and the sign of the first order leading principal minor. We refer to (a) as the 'cube approach' and (b) as the 'string approach'.

In table 3, performance-statistics are evaluated at (i) the N = 26 observed price points, denoted as $\boldsymbol{\psi}_{Ng}$, (ii) the C = 4 out of sample forecasts, $\boldsymbol{\psi}_{Cg}$ and (iii) the 8000 grid points $\boldsymbol{\psi}^{\square}_{g}$.

The first two estimation methods, 'local regularity' and 'global regularity', serve as a reference to the more interesting numerical results of the last three columns, in which comparisons between imposing the regularity conditions on $\boldsymbol{\psi}^{\square}_{g}$ versus imposing the regularity conditions on $\boldsymbol{\psi}^{string}_{g}$ are provided: The main result is that *the model fit measures are significantly improved, favoring the string approach, which suggests that it is worth reducing the size of* $\boldsymbol{\psi}$. Reductions in approximation errors can be achieved of over 40% and 83% for the AIM(1) and AIM(2), respectively. Further details on these percentages are presented in the last column.

We also supply performance statistics for the string approach evaluated over the cube grid $\boldsymbol{\psi}^{\square}_{g}$. We do not necessarily advocate such an approach (i.e. defining $\boldsymbol{\psi}$ on a subset of the region where subsequent inferences will be drawn). We rather include these results[31] to again emphasize the trade off between flexibility and regularity: The regional regularity approach can become useless when $\boldsymbol{\psi}$ does not cover the empirically relevant region (because it is likely that outside of $\boldsymbol{\psi}$

[30] The values of these 4 prices together with the 26 data points are provided in appendix.

regularity will be violated as is the case for AIM(1) and AIM(2)). This example underscores the advisability of considering the definition 1 carefully. In particular it is to be assumed that it is known prior to the estimation at which ranges of the data the model shall generate forecasts. Then we argue that, once it is ensured that the empirically relevant price set is regular, it is not particularly important if the function is irregular immediately outside the boundary of ψ because inferences will not be drawn from those regions.

5. Conclusion

In this paper we have developed a procedure for estimating functions subject to regional regularity conditions. Our method leads to improved model fit, is also computationally much faster and more efficient than previous approaches while imposing both curvature and monotonicity on the whole empirically relevant region of the regressor space. In fact its general nature presents a new method for the broader problem of estimating regression functions subject to nonlinear inequality constraints.

Our numerical examples illustrate that the tractability of the estimation procedure can be significantly improved by reducing the number of regularity checks. We hope this will lead to increased adoption of these faster simulation techniques in the applied literature. Another objective was to improve the in- and out-of-sample forecasts. The theoretical and numerical results provide evidence that by a) using the mode and/or by b) allowing ψ to be some connected non-convex set, the model fit statistics significantly improve. We further contributed to the literature by demonstrating that the commonly used Metropolis Hastings technique suffers from a bias of posterior density values. Finally we demonstrated that the commonly used posterior mean may be inappropriate as a point estimate. For both problems we suggested simple consistent alternatives.

---

[31] It is also interesting to see that even though the model fit statistics of the 'string approach' are clearly superior to the 'cube approach' when evaluated on ψ^{N}_{g}, this is not necessarily true when evaluated over the cubic region $\psi^{\square}_{g}$, (i.e. in the case of the AIM(2) the change in approximation errors are negative). The demand quantities for the out of sample prices in $\psi^{\square} \setminus \psi^{string}$ are calculated by (7).

Appendix: Proof of propositions and further explanations

Before we prove the cases outlined in table 1 we need to introduce two further set definitions. (1) For any *given* MCMC outcome $\mathbf{b}^{(*)} \in \Theta$, π can always be partitioned into two disjoint subsets, $\pi^{R}|\mathbf{b}^{(*)} \cup \pi^{IR}|\mathbf{b}^{(*)} = \pi$. We say that $f(\mathbf{p};\mathbf{b}^{(*)})$ is well behaved on the regular price set $\pi^{R}|\mathbf{b}^{(*)} = \{\mathbf{p} : \mathbf{i}(\mathbf{p}; \mathbf{b}^{(*)}) \geq \mathbf{0} \ \forall \ \mathbf{p} \in \pi\}$. (2) Since we are particularly interested in the behavior of the function within the set ψ, let us define $\psi^{R} = \psi^{R}|\mathbf{b}^{(*)} = \{\mathbf{p} : \mathbf{i}(\mathbf{p}; \mathbf{b}^{(*)}) \geq \mathbf{0} \ \forall \ \mathbf{p} \in \psi\} \subset \pi^{R}|\mathbf{b}^{(*)}$. It has the following features: If $f(\mathbf{p};\mathbf{b}^{(*)})$ is regular $\forall \ \mathbf{p} \in \psi$, then $\psi^{R} = \psi$. In general however $\psi^{R} \subset \psi$. For propositions 1a) to 2b) and 4, we prove sufficiency by contrapositive. To prove necessity is trivial and is omitted.

Proposition 1a:

$$\textit{Suppose} \left\{ \begin{array}{lll} \partial i_h/\partial p_1 \geq 0 \ \forall \ \mathbf{p} \in \psi \ \{\text{or } \partial i_h/\partial p_1 \leq 0 \ \forall \ \mathbf{p} \in \psi\} \\ \partial i_h/\partial p_2 \geq 0 \ \forall \ \mathbf{p} \in \psi \ \{\text{or } \partial i_h/\partial p_2 \leq 0 \ \forall \ \mathbf{p} \in \psi\} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \partial i_h/\partial p_K \geq 0 \ \ \forall \ \mathbf{p} \in \psi \ \{\text{or } \partial i_h/\partial p_2 \leq 0 \ \forall \ \mathbf{p} \in \psi\} \end{array} \right\} \quad (\textit{Property} \text{ I holds})$$

Iff $\mathbf{B} \subset \psi_h^{R}$, *then* $\psi_h^{R} = \psi$.

Proof of Proposition 1a[32]: Suppose not, then $\exists \ \mathbf{p}^* \in \psi^{IR} \backslash \mathbf{B}$ with $i_h(\mathbf{p}^*) < 0$. Further $\exists \ \mathbf{p}^{B} = [\, p_1^{B}, p_2^{B}, ..., p_K^{B} \,]^{T} \in \mathbf{B}$ which has the following property:

$$\begin{array}{ll} p_1^{B} \leq p_1^{*} \ \{\text{or } p_1^{B} \geq p_1^{*}\} \\ p_2^{B} \leq p_2^{*} \ \{\text{or } p_2^{B} \geq p_2^{*}\} \\ \vdots \qquad\qquad \vdots \\ \vdots \qquad\qquad \vdots \\ p_K^{B} \leq p_K^{*} \ \{\text{or } p_K^{B} \geq p_K^{*}\} \end{array}$$

[32] The 'or statements in the parenthesis {}' of *property* I are to be read as follows: in each k^{th} row either the statement without parenthesis or the statement within the parenthesis is true, except for the case that the derivative is zero on ψ. We explicitly allow that the signs across the K derivatives may be different. In the proof it then applies, that whenever in the k^{th} row of *property* I the derivative is nonnegative, then in the k^{th} row $p_k^{B} \leq p_k^{*}$. And equivalently, for nonpositive derivatives it applies $p_k^{B} \geq p_k^{*}$.

From *property* I it follows that $i_h(\mathbf{p}^B) \leq i_h(\mathbf{p}^*)$. Finally, since $i_h(\mathbf{p}^B) \leq i_h(\mathbf{p}^*) < 0$ it follows that $\mathbf{p}^B \in \boldsymbol{\psi}_h^{IR}$. Hence $\mathbf{B} \not\subset \boldsymbol{\psi}_h^R$. ***Q.E.D.***

We conclude that under the condition *property* I only bd($\boldsymbol{\psi}$) has to be evaluated. In practice however we cannot check for the connected set but approximate it by $\mathbf{B}_g$, thus still running the same risk as was criticized for the local regularity approach. Fortunately however, in many applications we can apply the results of the following proposition.

***Proposition 1b*:** *Suppose property* I *and property* II *hold. Iff* $\mathbf{z} = [\, p_1^{\min\{\max\}}, p_2^{\min\{\max\}}, \ldots, p_K^{\min\{\max\}}\,]^T \in \boldsymbol{\psi}_h^R$, *then* $\boldsymbol{\psi}_h^R = \boldsymbol{\psi}$.

***Proof of Proposition 1b*:** Suppose not, then $\exists\, \mathbf{p}^* \in \boldsymbol{\psi}^{IR} \backslash \{\mathbf{z}\}$ with $i_h(\mathbf{p}^*) < 0$ and by *property* I (see proposition 1a) $\exists\, \mathbf{p}^B \in \mathbf{B}$ with $i_h(\mathbf{p}^B) \leq i_h(\mathbf{p}^*)$, hence $\mathbf{p}^B \in \mathbf{B}^{IR}$. From *property* II it follows that $\exists$ one vertex point $\mathbf{z} = [z_1, z_2, \ldots, z_K]^T$ with the following property:

$$
\begin{aligned}
& z_1 \leq p_1^B \ \{\text{or } z_1 \geq p_1^B\} \\
& z_2 \leq p_2^B \ \{\text{or } z_2 \geq p_2^B\} \\
& \vdots \qquad\qquad \vdots \\
& \vdots \qquad\qquad \vdots \\
& z_K \leq p_K^B \ \{\text{or } z_K \geq p_K^B\}
\end{aligned}
$$

Hence $i_h(\mathbf{z}) \leq i_h(\mathbf{p}^B) \leq i_h(\mathbf{p}^*) < 0$. So $\mathbf{z} \in \boldsymbol{\psi}_h^{IR}$. ***Q.E.D.***

Since – under the conditions *property* I and *property* II – whenever $[\, p_1^{\min\{\max\}}, p_2^{\min\{\max\}}, \ldots, p_K^{\min\{\max\}}\,]^T \in \boldsymbol{\psi}_h^R$, then $\boldsymbol{\psi}_h^R = \boldsymbol{\psi}$, we conclude that only this single vertex point has to be checked.[33] For some inequality constraints the conditions of *property* I may however not hold. In that case the following result further greatly simplifies the Accept-Reject Algorithm.

[33] In case $\boldsymbol{\psi}$ is defined as the union of I $\boldsymbol{\psi}_i$, then the sum of vertices $[\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_I]$ are to be checked. Also note again, that for the k^{th} price we us $p_k^{\min}$ whenever in *property* I the k^{th} derivative is nonnegative and $p_k^{\{\max\}}$ whenever the k^{th} derivative is nonpositive.

***Proposition 2a*:** *Suppose* $\partial i_h/\partial p_m \geq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$ {or $\partial i_h/\partial p_m \leq 0 \;\; \forall \; \mathbf{p} \in \boldsymbol{\psi}$ } *and* $\partial i_h/\partial p_{-\mathrm{m}}$ *can take any value* (*property* III). *Iff* $\mathbf{B} \subset \boldsymbol{\psi}_h^{\mathrm{R}}$, *then* $\boldsymbol{\psi}_h^{\mathrm{R}} = \boldsymbol{\psi}$.

For the proof we need the following notation: Partition the $K \times 1$ vector $\mathbf{p}^* \in \boldsymbol{\psi}$ into the singular p_m^* and the K -1 $\times$ 1 vector $\mathbf{p}_{-m}^*$ and similarly partition $\mathbf{p}^{\mathrm{B}} \in \mathbf{B}$ into p_m^{B} and $\mathbf{p}_{-m}^{\mathrm{B}}$.

***Proof of Proposition 2a*:** Suppose not, then $\exists \; \mathbf{p}^* \in \boldsymbol{\psi}^{\mathrm{IR}} \backslash \{\mathbf{B}\}$ with $i_h(\mathbf{p}^*) < 0$. Further $\exists \; \mathbf{p}^{\mathrm{B}} = [\, p_1^{\mathrm{B}}, p_2^{\mathrm{B}}, ..., p_K^{\mathrm{B}} \,]^{\mathsf{T}} \in \mathbf{B}$ which has the following property:

$$p_m^{\mathrm{B}} \leq p_m^* \quad \{\text{or } p_m^{\mathrm{B}} \geq p_m^* \}$$

$$\mathbf{p}_{-m}^{\mathrm{B}} = \mathbf{p}_{-m}^*$$

By *property* III it follows that $i_h(\mathbf{p}^{\mathrm{B}}) = i_h(\, p_m^{\mathrm{B}}, \mathbf{p}_{-m}^{\mathrm{B}}) \leq i_h(\, p_m^*, \mathbf{p}_{-m}^*) = i_h(\mathbf{p}^*) < 0$. Hence $\mathbf{B} \not\subset \boldsymbol{\psi}_h^{\mathrm{R}}$.

Q.E.D.

Note that the assumptions of *property* III are much weaker than of *property* I and will hold for a wide set of common flexible functional forms and their respective ICFs, in which case we can omit checking the interior of $\boldsymbol{\psi}$. Similarly to proposition 1b, the following will further facilitate to the speed of MHARA.

***Corollary 2b*:** *Fix the* m^{th} *price axis from property III. Let* $\mathbf{S} \subset \mathbf{B} \subset \boldsymbol{\psi}$ *be that side of the hyperrectangle*, *which is orthogonal to the* m^{th} *price-axis and for which* $p_m^{\mathrm{S}} = p_m^{\min\{\max\}} \;\; \forall$ $(\, p_m^{\mathrm{S}}, \mathbf{p}_{-m}^{\mathrm{S}}) \in \mathbf{S}$. *Suppose property* II *and property* III *hold. Iff* $\mathbf{S} \subset \boldsymbol{\psi}_h^{\mathrm{R}}$, *then* $\boldsymbol{\psi}_h^{\mathrm{R}} = \boldsymbol{\psi}$.

Proof of Corollary 2b*:** The proof follows the same logic as the proof of proposition 1b. ***Q.E.D.

In other words, if *property* II and III hold, then it is only necessary to evaluate **S** which is that side of the hyperrectangle which is orthogonal to the m^{th} price-axis and on which the value of p_m is either a) smallest, in the case that $\partial i_h/\partial p_m \geq 0$ or b) largest, in the case that $\partial i_h/\partial p_i \leq 0$. For illustration, see fig. A1.

-- INSERT FIG. A1 --

The following proposition defines conditions under which it is sufficient to check the vertices of a convex set $\boldsymbol{\psi}$ only, or in case of a hypercube, its 2^K vertices $\mathbf{Z}_h$ [34].

Proposition 3*: *Suppose property IV holds. Iff $\mathbf{Z}_h \in \boldsymbol{\psi}_h^R$, *then* $\boldsymbol{\psi}_h^R = \boldsymbol{\psi}$.

Proof of Proposition 3*: Lemma**: A quasi-concave function i_h has the property that its upper contour set $\mathbf{U}_\omega = \{\mathbf{p}: i_h \geq \omega, \mathbf{p} \in \boldsymbol{\psi}, \omega \in \mathbb{R}^1\}$ is convex. Let $\boldsymbol{\psi}_h^R = \mathbf{U}_0 = \{\mathbf{p}: i_h \geq 0, \mathbf{p} \in \boldsymbol{\psi}\}$ be the upper contour set $\mathbf{U}_0$ evaluated at $\omega = i_h = 0$. Since $\boldsymbol{\psi}$ is convex by *property* IV it follows that $\boldsymbol{\psi}_h^R = \mathbf{U}_0 \cap \boldsymbol{\psi}$ is convex (since the intersection of convex sets is convex). Further, since any convex set is connected and $\boldsymbol{\psi}_h^R$ includes the vertices $\mathbf{Z}_h$ of $\boldsymbol{\psi}$, $\boldsymbol{\psi}_h^R = \boldsymbol{\psi}$. ***Q.E.D.

Remark: In order to identify quasiconcavity of property IV, in practice it is useful to make use of the bordered Hessians of $\mathbf{i}(\cdot)$, see e.g. SIMON and BLUME (pp.523-531:1994).

***Proposition 4*:** *Suppose the regularity conditions to be imposed belong to a subset of the following properties: (a) nonpositve slope, (b) nonnegative slope, (c) convexity, or (d) concavity. Suppose property V holds. Iff* $\mathbf{S}^* \in \boldsymbol{\psi}^R$ *then* $\boldsymbol{\psi}^R = \boldsymbol{\psi}$.

***Proof of Proposition 4*:** Suppose not, then $\exists\ \mathbf{p}^* \in \boldsymbol{\psi}^{IR} \backslash \mathbf{S}^*$ for which either (a) nonpositve slope, (b) nonnegative slope, (c) convexity, or (d) concavity is violated.

First suppose monotonicity, (a) or (b), is violated at $\mathbf{p}^*$. Then at least one element $\partial f(\mathbf{p}^*)/\partial p_k$ of the $K \times 1$ gradient vector $\partial f(\mathbf{p}^*)/\partial \mathbf{p}$ is wrong in sign. By the property of a homogenous of degree α function, $\alpha \in \mathbb{R}^1$, we have $\partial f(t\mathbf{p}^*)/\partial \mathbf{p} = t^{\alpha-1}\partial f(\mathbf{p}^*)/\partial \mathbf{p}\ \forall\ t > 0$. This implies that the signs of the elements of the gradient vector evaluated at $t\mathbf{p}^*$ do not change to the gradient vector evaluated at $\mathbf{p}^*$, and hence any $t\mathbf{p}^*$ is irregular as well. Consequently, also irregular is the point $\mathbf{p}^{\mathbf{S}^*} \in \mathbf{S}^* \cap l(\mathbf{0},\mathbf{p}^*)$ at which the ray through the origin and $\mathbf{p}^*$ intersects with shield $\mathbf{S}^*$.

[34] If $\boldsymbol{\psi}_i$ is defined as a part of a hyperplane in $\boldsymbol{\pi}$, the number of vertices might be different from 2^K. For example, in the case that $\boldsymbol{\psi}_i$ has the form of a line, we just have two instead of 2^K vertices, the starting and the ending point of the line.

Now suppose curvature, (c) or (d), is violated at $\mathbf{p}^*$. Then the Hessian evaluated at $\mathbf{p}^*$, $\mathbf{H}|_{\mathbf{p}^*}$, does not maintain the correct semi-definiteness. Again, by the property of homogenous functions we have $\partial f^2(t\mathbf{p}^*)/\partial\mathbf{p}\partial\mathbf{p}' = t^{\alpha-2}\partial^2 f(\mathbf{p}^*;\beta)/\partial\mathbf{p}\partial\mathbf{p}' \; \forall \; t > 0$. Since $\mathbf{H}|_{t\mathbf{p}^*}$ only differs from $\mathbf{H}|_{\mathbf{p}^*}$ by the multiple $t^{\alpha-2}$ the definiteness of the matrices is identical, hence $t\mathbf{p}^* \in \boldsymbol{\psi}^{IR} \; \forall \; t > 0$. Consequently, also irregular is the point $\mathbf{p}^{S^*} \in \mathbf{S}^* \cap l(\mathbf{0},\mathbf{p}^*)$. ***Q.E.D.***

Appendix of figures and tables

Fig 1: Irregular function

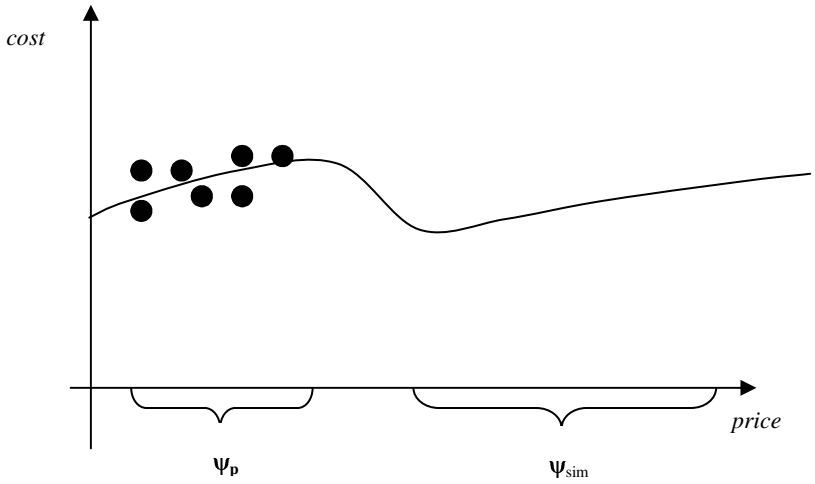


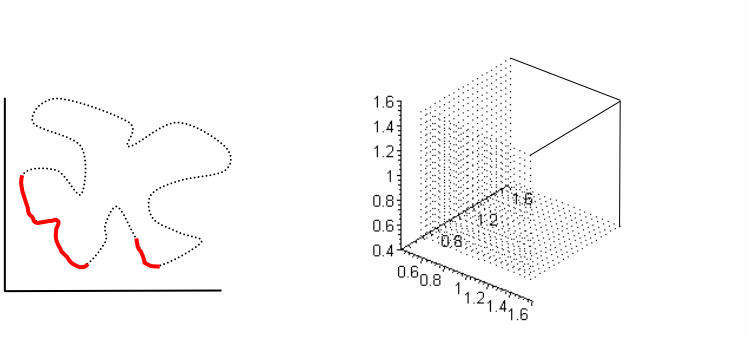
Fig. 1 depicts an example where $\psi_{\mathbf{p}}$ includes all observed data points (each dot represents an observed (cost, price) combination used for estimating the cost function), and ψ_{sim} includes the region at which inferences will be drawn for simulation purposes. However, $\psi = \psi_{\mathbf{p}} \cup \psi_{sim}$ violates the requirement that it is one connected set. The graph shows that imposing concavity and monotonicity at both regions $\psi_{\mathbf{p}}$ and ψ_{sim} does not necessarily generate overall regularity and can lead to spurious forecasts because costs must not decline with rising input prices.

Table 1: Sufficient conditions for defining the evaluation set as a subsets of ψ

Case	*Property* I	*Property* II	*Property* III	*Property* IV	*Property* V	ψ_h	Support generated by the h^{th} grid	Proposition
1	+					boundary $\mathbf{B}_h$	$\Theta^R\|\mathbf{B}_{gh} \supset \Theta^R\|\psi$	1a
2	+	+				one vertex $\mathbf{z}_h$	$\Theta^R\|\mathbf{z}_h = \Theta^R\|\psi$	1b
3			+			boundary $\mathbf{B}_h$	$\Theta^R\|\mathbf{B}_{gh} \supset \Theta^R\|\psi$	2a
4		+	+			side $\mathbf{S}_h$	$\Theta^R\|\mathbf{S}_{gh} \supset \Theta^R\|\psi$	2b
5				+		all vertices $\mathbf{Z}_h$	$\Theta^R\|\mathbf{Z}_h = \Theta^R\|\psi$	3
6					+	shield $\mathbf{S}^*$	$\Theta^R\|\mathbf{S}^* \supset \Theta^R\|\psi$	4

Symbol ψ_h is a placeholder for $\mathbf{B}_h$, $\mathbf{S}_h$, $\mathbf{S}^*$, $\mathbf{z}_h$, and $\mathbf{Z}_h$. For the proofs of the statements in the table see section A1 of the appendix.

Fig. 2: Illustrations of evaluation grids for the Accept-Reject algorithm



To the left, an example of a shield $\mathbf{S}^* \subset \psi$ is displayed. To the right the shield grid $\mathbf{S}_g^* \subset \psi = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\}$ which we also use for the second principal minor test for the AIM(2) in section 4.

Fig. 3: Violations on the price grid $\psi^{\square}_g$ in the case of the local regularity approach

In 19.09% of the grid points monotonicity is violated (left cube) and in 3.11% concavity is violated (right cube). Each black dot is one grid point where violation occurs.

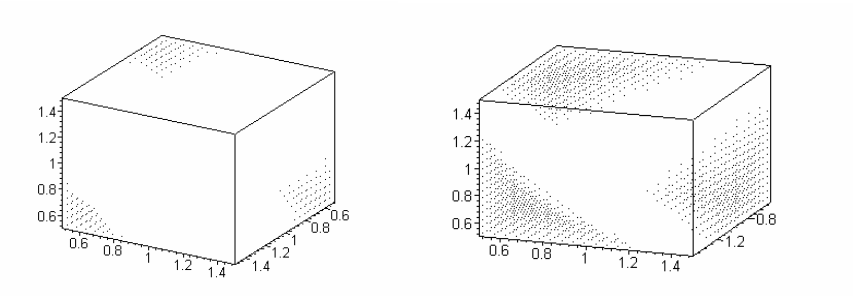


Table 2: Global, regional and local approach - comparison based on AIM cost functions[1]

Model	Forecast Error and Regularity Violations evaluated over $\psi^{\square}_{g}$	Estimation Approach			
		Local Regularity[2]	Global Regularity[2]	Regional Regularity	
				Mean	Mode
AIM(1)	AAAE	0.05208	0.14395	0.095523	0.093291
	MAE	-0.19692	0.469	0.29045	0.28540
	Concavity Violations	0%	0%	0%	0%
	Monotonicity Violations	17.33%	0%	0%	0%
AIM(2)	AAAE	0.02056	0.13266	0.040248	0.036739
	MAE	-0.07563	0.40808	0.11591	0.10759
	Concavity Violations	3.11%	0%	0%	0%
	Monotonicity Violations	19.09%	0%	0%	0%

(1) Experiment based on table 1 and table 2 of Terrell (1995): True data generation process: CES technology with parameter settings $a_i = 1$; $\rho = 0.75$. In order to provide a benchmark for the average and largest error, the CES-input demand data $\mathbf{x}_k$ have, as in TERRELL (1995), mean of $8000^{-1} \sum_{g=1}^{8000} x_{gk} = 0.2552 \ \forall \ k$ and standard deviation of std($\mathbf{x}_k$) = 0.2230 $\forall$ k over the evaluation grid $\psi^{\square}_{g}$.

(2) Some considerable differences exist between our and TERRELL's (1995) results. (a) Local Regularity AIM(2): Instead of 3.11% TERRELL found 1.6% of concavity violations. (b) He calculated error statistics in the column 'global approach' which are about 3-4 times higher for the AAAE and 1.5 times higher for the MAE than our results: AIM(1): AAAE = 0.64146, MAE = -0.84186; AIM(2): AAAE = 0.47073, MAE = -0.63968. After careful consideration, we believe that the results in our table are the correct ones.

Table 3: Local, global, regional cube and regional string approach - comparison based on AIM cost functions

Model	Model Performance Statistics		Estimation Approach												Percentage change in error statistics of the string approach relative to the cube approach		
			Local Regularity, imposed at $\mathbf{p}^M$			Gobal Regularity			Regional Regularity: imposed on $\psi^{\square}_g$ (cube approach)			Regional Regularity: imposed on ψ^{string}_g (string approach)					
	Forecast Error / Regularity Violations	evaluated at	Input 1	Input 2	Input 3	Input 1	Input 2	Input 3	Input 1	Input 2	Input 3	Input 1	Input 2	Input 3	Input 1	Input 2	Input 3
AIM(1)	AAAE	ψ^N_g	0.0316	0.0363	0.0174	0.1655	0.1521	0.1416	0.1008	0.0992	0.1080	0.0358	0.0394	0.0209	64.54%	60.27%	80.63%
	MAE		0.0953	0.0909	-0.0477	0.4199	0.4591	0.4885	0.1906	0.2037	0.4243	0.1056	0.1217	-0.0622	44.62%	40.27%	85.33%
	Concavity Violations			0.00%			0.00%			0.00%			0.00%				
	Monotonicity Violations			11.54%			0.00%			0.00%			0.00%				
	AAAE	ψ^C_g	0.0095	0.0181	0.0174	0.1118	0.1513	0.1102	0.0794	0.1006	0.1121	0.0143	0.0313	0.0277	81.97%	68.91%	75.27%
	MAE		-0.0192	0.0502	0.0326	-0.1944	0.4220	0.2526	-0.1294	0.2037	0.2487	-0.0284	0.0888	-0.0538	78.01%	56.42%	78.34%
	Concavity Violations			0.00%			0.00%			0.00%			0.00%				
	Monotonicity Violations			0.00%			0.00%			0.00%			0.00%				
	AAAE	$\psi^{\square}_g$	0.0467	0.0472	0.0484	0.1484	0.1447	0.1425	0.0971	0.0952	0.1094	0.0483	0.0491	0.0467	50.26%	48.40%	57.35%
	MAE		-0.2797	-0.2886	-0.2963	0.4202	0.4594	0.5360	-0.2920	-0.2843	0.4357	-0.2734	-0.2560	-0.2698	6.37%	9.93%	38.08%
	Concavity Violations			0.00%			0.00%			0.00%			0.00%				
	Monotonicity Violations			32.66%			0.00%			0.00%			28.01%				
AIM(2)	AAAE	ψ^N_g	0.0042	0.0039	0.0025	0.1514	0.1382	0.1299	0.0470	0.0475	0.0459	0.0070	0.0082	0.0055	84.99%	82.83%	88.02%
	MAE		-0.0165	-0.0115	0.0111	0.3838	0.4051	0.4167	0.1256	0.1185	0.1259	-0.0199	-0.0272	-0.0192	84.16%	77.02%	84.78%
	Concavity Violations			15.39%			0.00%			0.00%			0.00%				
	Monotonicity Violations			0.00%			0.00%			0.00%			0.00%				
	AAAE	ψ^C_g	0.0024	0.0028	0.0035	0.0962	0.1353	0.0936	0.0286	0.0465	0.0527	0.0013	0.0024	0.0024	95.49%	94.90%	95.52%
	MAE		-0.0078	-0.0093	0.0111	-0.1764	0.3911	0.2189	-0.0522	0.0992	0.1110	-0.0020	0.0033	-0.0028	96.26%	96.64%	97.44%
	Concavity Violations			25.00%			0.00%			0.00%			0.00%				
	Monotonicity Violations			0.00%			0.00%			0.00%			0.00%				
	AAAE	$\psi^{\square}_g$	0.0142	0.0151	0.0133	0.1369	0.1329	0.1296	0.0470	0.0470	0.0432	0.0153	0.0154	0.0155	67.54%	67.12%	64.18%
	MAE		0.3782	0.4073	-0.3278	0.3865	0.4082	0.4350	0.1391	0.1298	0.1287	0.1459	0.1335	0.1044	-4.93%	-2.85%	18.86%
	Concavity Violations			26.46%			0.00%			0.00%			11.30%				
	Monotonicity Violations			0.69%			0.00%			0.00%			9.90%				

Simulation experiment based on table 1 and table 2 of Terrell (1995): True Data Generation Process: CES technology with parameter settings $a_i = 1; \rho = 0.75$.

Fig. 4: The String grid ψ^{string}_{g}

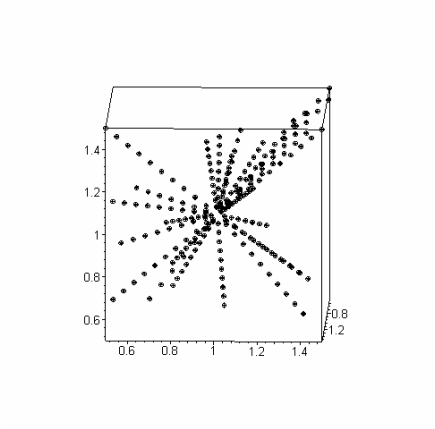
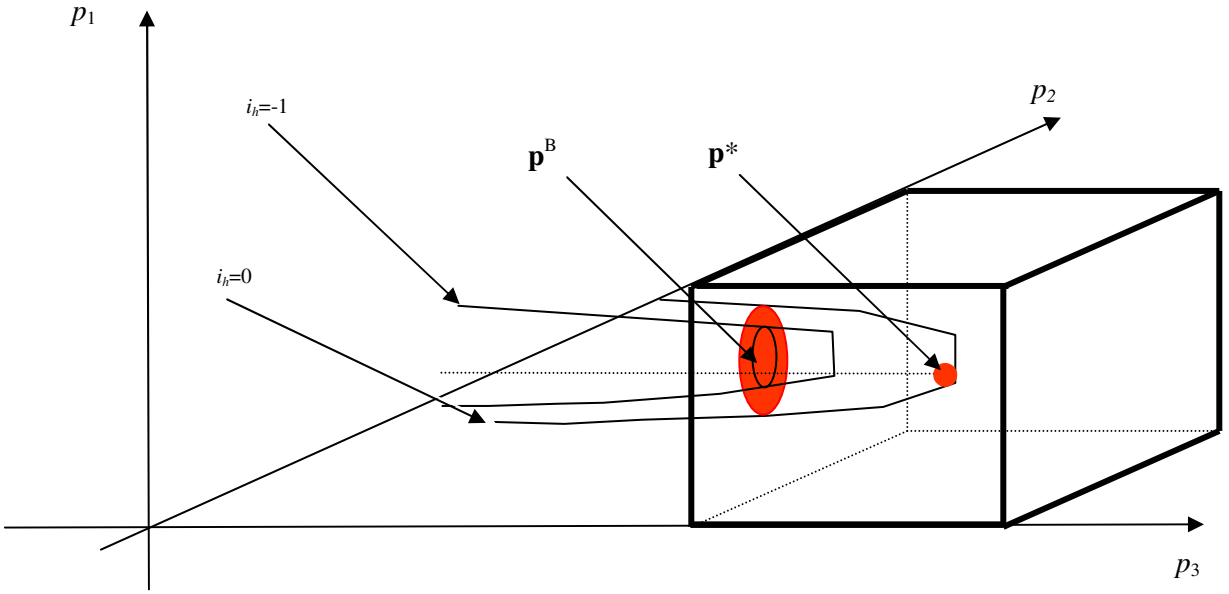


Fig A1: ICF level sets i_h = -1 and i_h = 0 in price space π

If *property* II and *property* III hold, **p*** is irregular, and $\partial i_h/\partial p_3 \geq 0$, then the boundary side **S** facing towards the p_1–p_2 level contains irregular points $\mathbf{p}^B \in \mathbf{S}^{IR} \subset \mathbf{S}$ as well. $\mathbf{S}^{IR}$ is shaded in red. The set $\psi \subset \pi$ is indicated by the cube.



A2: Input price observations and out of sample points used for experiment II

26×3 Input price observation matrix **P**

n	input price 1	input price 2	Input price 3
1	0.59404	0.56000	0.55000
2	0.52200	0.68344	0.84049
3	0.55812	1.05000	1.18890
4	0.57451	1.49900	1.46040
5	0.94357	0.54122	0.81883
6	0.69551	0.78415	0.60475
7	0.82898	0.78613	0.73893
8	0.84189	1.15940	1.09310
9	0.80024	1.49740	1.45910
10	1.12530	0.56597	1.08850
11	1.15600	0.95502	1.37150
12	1.38970	1.04470	0.64871
13	1.21790	1.38860	0.76997
14	1.02370	1.21050	1.34420
15	1.09690	1.44260	1.47270
16	1.46630	0.58908	1.30410
17	1.44160	1.02990	1.41120
18	1.41350	1.14770	1.47790
19	1.38970	1.41070	0.61131
20	1.48110	1.43560	0.79465
21	1.48060	1.34620	1.06060
22	1.43460	1.42840	1.46580
23	0.50000	0.50000	0.50000
24	1.50000	1.50000	1.50000
25	1.50000	0.50000	1.50000
26	0.50000	1.50000	1.50000

$C =$ 4 scenario input price vectors

c	input price 1	input price 2	input price 3
1	1.00000	1.00000	1.00000
2	1.28870	1.26140	0.87679
3	3.00000	3.00000	3.00000
4	4.39890	1.76720	3.91230

References

ADKINS, LEE C., DAN S. RICKMAN and ABID HAMEED (2002): Bayesian Estimation of Regional Production for CGE Modeling. Paper presented at the fourteenth International Conference on Input-Output Techniques October 10-15, 2002, Montréal, Canada.

BARNETT, WILLIAM A. (1976): Maximum Likelihood and Iterated Aitken Estimation of Non-Linear Systems of Equations. *Journal of the American Statistical Association*, 71, 354-360.

BARNETT, WILLIAM A. (1985): The Minflex-Laurent Translog Flexible Functional Form. *Journal of Econometrics*, 30, 33-44.

BARNETT, WILLIAM A., JOHN GEWEKE, and MICHAEL WOLFE (1991): Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model. *Journal of Econometrics*, 49, No. 1/2, pp.5-50.

BARNETT, WILLIAM A. (2002): Tastes and technology: curvature is not sufficient for regularity. *Journal of Econometrics*, 108, 199-202.

BARNETT, WILLIAM A. and MEENAKSHI PASUPATHY (2003): Regularity of the Generalized Quadratic Production Model: A Counterexample. *Econometric Reviews*, 22-2, pp. 135-154.

CHIB, S. and GREENBERG, E. (1996): Markov Chain Monte Carlo Methods in econometrics. *Econometric Theory*, 12, pp.409-431.

CHEN, M.H., Q.M. SHAO and J.G. IBRAHIM (2000): Monte Carlo Methods in Bayesian Computation. Springer, New York.

CHUA, C.L, W.E. GRIFFITHS, and C.J. O'DONNELL (2001): Bayesian Model Averaging in Consumer Demand Systems With Inequality Constraints. *Canadian Journal of Agricultural Economics*, 49(2001): 269-291.

CUESTA, RAFAEL A., CHRISTOPHER J. O'DONNELL, TIMOTHY J. COELLI and SATBIR SINGH (2001): Imposing Curvature Conditions on a Production Frontier: With Applications to Indian Dairy Processing Plants. *CEPA Working Papers*, No. 2/2001, ISBN 1 86389 749 6, School of Economics, University of New England, Armidale.

DIEWERT, W. E., and WALES, T. J. (1987): Flexible Functional Forms and Global Curvature Conditions. *Econometrica* 55, 43-68.

DRISCOLL, PAUL J. and RICHARD BOISVERT (1991): Dual Second- and Third-Order Translog Models of Production, *American Journal of Agricultural Economics*, 73, pp.1146-1160.

GALLANT, A. RONALD and GENE H. GOLUB (1984): Imposing Curvature Restrictions on Flexible Functional Forms. *Journal of Econometrics*, 26, pp.295-322.

GRIFFITHS, W.E., C.L. SKEELS and D. CHOTIKAPANICH (2002): Sample Size Requirements for Estimation in SUR Models. In A. ULLAH, A., CHATURVEDI and A. WAN, Eds., Handbook of Applied Econometrics and Statistical Inference, New York: Marcel Dekker.

GRIFFITHS, WILLIAM E., C.J. O'DONNELL and A. TAN-CRUZ (2000): Imposing Regularity Conditions on a System of Cost and Factor Share Equations. *Australian Journal of Agricultural and Resource Economics*, 44, pp.107-127.

GRIFFITHS, WILLIAM E. (2003): Bayesian Inference in the Seemingly Unrelated Regressions Model. In D.E.A. Giles, editor, Computer-Aided Econometrics, New York: Marcel Dekker, to appear, 2003.

IVALDI, MARC, NORBERT LADOUX, HERVÉ OSSARD and MICHEL SIMIONI (1996): Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers. *Journal of applied econometrics*, 11(6), pp.649-667.

JEHLE, GEOFFREY A. and PHILIP J. RENY (2000): Advanced Microeconomic Theory. Second Edition, Addison Wesley, Boston.

FISCHER, FLEISSIG and SERLETIS (2001): An Empirical Comparison of Flexible Demand System Functional Forms. *Journal of Applied Econometrics*, 16(1), pp.59-80.

FLEISSIG, ADRIAN R., TERRY KASTENS and DEK TERRELL: (1997): Semi-nonparametric Estimates of Substitution Elasticities. *Economic Letters*, 54, pp.209-215.

FLEISSIG, ADRIAN R., TERRY KASTENS and DEK TERRELL: (2000): Evaluating the Semi-nonparametric Fourier, AIM, and Neural Networks Cost Functions. *Economic Letters*, 68, pp.235-244.

KLEIT, ANDREW N. and DEK TERRELL (2001): Measuring Potential Efficiency Gains from Deregulation of Electricity Generation: A Bayesian Approach. *The Review of Economics and Statistics*, 83(3), pp.523-530.

KOOP, GARY, JACEK OSIEWALSKI and MARK FJ STEEL (1994): Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function. *Journal of Business and Economic Statistics*, 12(3) pp.339-46.

KOOP, GARY, JACEK OSIEWALSKI and MARK FJ STEEL (1997): Bayesian Efficiency Analysis through Individual Effects: Hospital cost frontiers. *Journal of Econometrics*, 76, pp.77-105.

LAU, LAWRENCE J. (1978): Testing and Imposing Monotonicity, Convexity, and Quasi-Convexity Constraints. In: Production Economics: A Dual Approach to Theory and Applications (volume 1), eds. FUSS, M and D. MCFADDEN, Amsterdam: North-Holland, 1978, pp.409-453.

LAU, LAWRENCE J. (1986): Functional Forms in Econometric Model Building. Chapter 26 in Handbook of Econometrics, Vol. 3, Ed. GRILICHES, ZVI and INTRILIGATOR, MICHAEL D., Elsevier Science, Amsterdam, North Holland, pp.1515-1566.

MAGNUS, J.R. (1979): Substitution between Energy and Non-Energy Inputs in the Netherlands 1950-1976. *International Economics Review*, 20, pp.465-484.

MAS-COLELL, ANDREAU, MICHAEL D. WHINSTON and JERRY R. GREEN (1995): Microeconomic Theory. Oxford University Press.

MATZKIN, ROSA L. (1994): Restrictions of Economic Theory in Nonparametric Methods. Handbook of Econometrics, Volume 4, chapter 42, Ed. ROBERT F. ENGLE and DANIEL L. MCFADDEN, pp. 2524-58.

O'DONNELL, CHRISTOPHER J., C. RICHARD SHUMWAY and V. ELDON BALL (1999): Input Demands and Inefficiency in U.S. Agriculture. *American Journal of Agricultural Economics*, 81, November, pp.865-880.

O'DONNELL, C.J., A.N. RAMBALDI and H.E. DORAN (2001): Estimating Economic Relationships Subject to Firm- and Time-Varying Equality and Inequality Constraints. *Journal of Applied Econometrics*, 16(6):709-726.

O'DONNELL, C.J., and T. COELLI (2003): A Bayesian Approach to Imposing Curvature on Distance Functions. Paper presented at the Australasian Meeting of the Econometric Society, Sydney 2003.

RYAN, DAVID L. and TERENCE J. WALES (1998): A Simple Method for Imposing Local Curvature in Some Flexible Consumer Demand Systems. *Journal of Business and Economic Statistics*, 16(3), pp.331-38.

SIMON, CARL P. and LAWRENCE BLUME (1994): Mathematics for Economists. W.W. Norton, New York.

TERRELL, D. (1995): Flexibility and Regularity Properties of the Asymptotic Ideal Production Model. *Econometric Reviews* 14(1), pp.1-17.

TERRELL, D. (1996): Incorporating monotonicity and concavity conditions in flexible functional forms, *Journal of Applied Econometrics* 11, pp.179-194.

TERRELL, D. and IMAD DASHTI (1997): Imposing Monotonocity and Concavity Restrictions on Stochastic Frontiers. Working Paper, Department of Economics, Louisiana State University, E. J. Ourso College of Business Administration.

WOLFF, H., T. HECKELEI, and R.C. MITTELHAMMER (2003): Imposing Regional Regularity on Flexible Functional Forms. Technical Working Paper 110 of the International Marketing Program for Agricultural Commodities & Trade (IMPACT), Washington State University, Pullman.

ZELLNER, A. (1971): An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, New York.