Comput Econ (2010) 36:309–339
DOI 10.1007/s10614-010-9215-1

Imposing Curvature and Monotonicity on Flexible Functional Forms: An Efficient Regional Approach

Hendrik Wolff · Thomas Heckelei · Ron C. Mittelhammer

Accepted: 20 April 2010 / Published online: 14 May 2010

Abstract In many areas of economic analysis, economic theory restricts the shape of functions. Examples are the monotonicity and curvature conditions that apply to utility, profit, and cost functions. Here we extend upon a currently available estimation method (Terrell, J Appl Econometr 11:179–194, 1996) for imposing regularity *regionally* on a connected subset of the regressor space. Our method offers important advantages 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 through higher flexibility, which generally leads to a better model fit to the sample data. Specific contributions of this paper are (a) to increase the computational speed, (b) to provide regularity preserving point estimates, and (c) to illustrate the benefits of this revised regional approach via numerical simulation results.

Keywords Nonlinear inequality constraints · Flexible functional forms · Metropolis-Hastings · Accept–Reject algorithm · Cost function · Regularity conditions

JEL Classification C51 · D21 · C11

H. Wolff (✉)
Department of Economics, University of Washington, 349 Savery Hall, Box 353330, Seattle, WA 98195-3330, USA
e-mail: hgwolff@u.washington.edu

T. Heckelei
Institute for Food and Resource Economics, University of Bonn, Bonn, Germany

R. C. Mittelhammer
School of Economic Sciences, Washington State University, Pullman, WA, USA

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 at one point in the regressor space. Recently, higher (than second) order series expansions, such as the Fourier and the Asymptotically Ideal Production Model (AIM), have been suggested (e.g. Gallant and Golub 1984; Barnett et al. 1991; Koop et al. 1994). These representations promise a better fit to the data as they transition from local to global flexibility and as the order of the expansion increases. Even more recently new promising nonparametric estimation techniques that account for shape restrictions (originally proposed by Hildreth, 1954) have garnered increasing attention in the literature (Matzkin 1994; Tripathi 2000; Aït-Sahalia and Duarte 2003; Racine and Parmeter 2008). The advantage of such an approach is that no assumption about a parametric functional form, or a series expansion thereof, has to be imposed. However, this advantage comes at the cost of lower asymptotic convergence rates as well as sometimes unknown asymptotic distributions. This is particularly a problem in higher dimensional settings, where not univariate constraints are imposed but regularity is required in multi-input or multi-output settings. In this paper we focus on the problem of the estimation of parametric functional forms.

Unfortunately, the estimated parametric functions that model economic behavior frequently violate curvature and monotonicity restrictions and the propensity for such violations can increase with the order of flexibility.[1] 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, p. 43).

[1] A number of authors (Diewert and Wales 1991; Salvanes and Tjøtta 1998; Barnett 2002; Barnett and Pasupathy 2003) summarize the literature on empirical productivity and supply and demand analysis noting that it has become customary to use flexible functional forms, but rarely have the shape conditions been formally tested and as a consequence the first order properties of duality theory can fail. For example, Diewert and Wales (1991) tested a popular AER paper by Evans and Heckman (1984) which estimated the cost function of the Bell System (in order to test whether the Bell System is a natural monopoly). Diewert and Wales (1991) express serious concerns whether this test was meaningful as it was based on a non-regular estimated functional form. In particular, Diewert and Wales (1991) found that the estimated function by Evans and Heckman exhibited negative marginal cost in most of the relevant regressor space and expressed serious concern as to whether their test was at all meaningful. Later Salvanes and Tjøtta (1998) proposed a procedure to calculate the range of the regressor values where the restrictions imposed by economic theory are satisfied. Barnett (2002) and Barnett and Pasupathy (2003) discuss two specific examples where the literature failed to estimate functions that are consistent with economic theory. The first example is concerned with the modeling of financial intermediaries and the second with estimating the technology of manufacturing firms as estimated by Barnett et al. (1995). Barnett (2002) provides striking visual examples of how wrong the estimation can be by displaying isoquants that are of highly implausible shape.

In this paper we propose and illustrate a Bayesian estimation procedure for imposing regularity conditions via nonlinear inequality constraints. Building upon work by Terrell (1996), the conditions are imposed on a connected[2] subset of the domain of the function being estimated. The connected subset represents what we refer to as the *empirically relevant region*, and is defined by the model analyst. This *regional approach* offers important advantages over the *local* approach by imposing theoretical consistency not only locally at a given evaluation point, but also over the entire 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 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 underscore 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[3] 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. 1552–1557) 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 maintain *local flexibility*), which is a fact that does not seem to be generally appreciated in the literature on globally flexible functional forms.[4]

[2] A connected set is such that any two points in the set can be connected by a continuous curve 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 $\mathbf{U} \cap \mathbf{V} = \emptyset$ and $\mathbf{U} \cup \mathbf{V} = \mathbf{X}$.

[3] 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.

[4] 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.

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,[5] that can also be rife with statistical testing/verification problems. In Gallant and Golub (1984) 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 Sect. 1.3.

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 (1984) discussed the possibility of imposing regularity conditions on a predefined regular region $\boldsymbol{\psi}$ 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 (1984) 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 (Terrell 1996) 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 $\boldsymbol{\psi}$ 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 $\boldsymbol{\beta}$, conditional on the endogenous

[5] 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 et al. (1991, p. 10) and more recently Terrell (1995); Terrell (1996), Ivaldi et al. (1996), Fleissig et al. (1997, 2000), Jensen (1997), Ryan and Wales (1998), Fischer et al. (2001) for studies evaluating these mentioned and other competing forms. We recommend Barnett et al. (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.

variable $\mathbf{y}$, $p_u(\boldsymbol{\beta}|\mathbf{y})$, is derived that does not incorporate the regularity conditions. Third, a Gibbs sampler is used to draw parameter vector outcomes from $p_u(\boldsymbol{\beta}|\mathbf{y})$, 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 $\boldsymbol{\Theta}$ of the unconstrained posterior $p_u(\boldsymbol{\beta}|\mathbf{y})$. However, this can be time consuming if, as is often the case in practice, the regular region is only a small subset of $\boldsymbol{\Theta}$ (Terrell 1996).

To overcome the latter problem, Griffiths et al. (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[6] 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 principal 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 $\boldsymbol{\psi}$ 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 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 $\boldsymbol{\psi}$ (objective I). If these conditions are satisfied, a twofold benefit results:
   (i) Imposition of regularity in $\boldsymbol{\psi}$ 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.
2. allowing $\boldsymbol{\psi}$ to be some connected non-convex set, which can significantly increase the model fit achievable from estimation (objective II).

[6] Literature on applications of MHARA include Koop et al. (1994), O'Donnell et al. (1999), Griffiths et al. (2000), Griffiths (2003), Chua et al. (2001), Cuesta et al. (2001), Kleit and Terrell (2001), O'Donnell et al. (2001) and O'Donnell and Coelli (2003).

3. 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 Sect. 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 methodological contributions. Examples using AIM functional forms are given in Sect. 4 in order to illustrate the methodology and demonstrate empirical relevance. A final section presents conclusions and the appendix contains all necessary proofs as well as additional details relating to the implementation of the estimation procedure.

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, $c(\mathbf{p}; \boldsymbol{\beta})$, whereby $\mathbf{p} = [p_1, p_2, \ldots, p_K]^T \in \boldsymbol{\pi}$ are K input prices, $\boldsymbol{\pi}$ denotes the orthant of strictly positive prices in $\Re^K$, and $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ is the parameter vector to be estimated. According to economic theory $c(\mathbf{p}; \boldsymbol{\beta})$ must be concave and nondecreasing in $\mathbf{p}$ (Mas-Colell et al. 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*, $\mathbf{i}\,(\mathbf{p}; \boldsymbol{\beta}) \equiv [i_1, i_2, \ldots, i_H] : (\boldsymbol{\pi} \times \boldsymbol{\Theta}) \to \Re^H$, whereby the restrictions hold whenever, for a given $\boldsymbol{\beta}$, $\mathbf{i}()$ is 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 $c(\mathbf{p}; \boldsymbol{\beta})$ is a twice continuously differentiable, linear homogenous in $\mathbf{p}$ unit cost function with $K = 2$ input prices, then the inequality constraints could be defined as[7]

$$i_1 = \partial c(\mathbf{p}; \boldsymbol{\beta})/\partial p_1, \quad i_2 = \partial c(\mathbf{p}; \boldsymbol{\beta})/\partial p_2, \quad i_3 = -\partial^2 c(\mathbf{p}; \boldsymbol{\beta})/\partial p_1^2 \quad \text{and}$$
$$i_4 = -\partial^2 c(\mathbf{p}; \boldsymbol{\beta})/\partial p_2^2$$

[7] Note that nonnegativity of i_1 and i_2 imposes monotonicity. Nonnegativity of $i_{\prime 3}$ and i_5 imposes negative semi-definiteness on the Hessian $\partial^2 f(\mathbf{p}; \boldsymbol{\beta})/\partial \mathbf{p} \partial \mathbf{p}\prime$. Since by linear homogeneity of $f(\cdot)$ the Hessian has rank $K - 1$, it is not necessary to generate an additional inequality constraint function to sign the Kth principal minor.

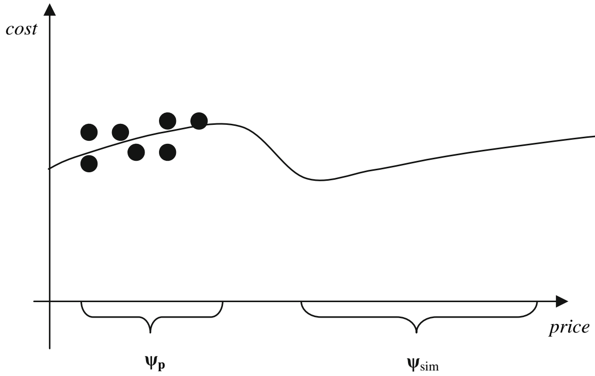


Fig. 1 Irregular function. Figure depicts an example where $\boldsymbol{\psi}_p$ includes all observed data points (each dot represents an observed (cost, price) combination used for estimating the cost function), and $\boldsymbol{\psi}_{sim}$ includes the region at which inferences will be drawn for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

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 set $\boldsymbol{\psi}$ is a closed[8] and connected subset of $\boldsymbol{\pi}$ that covers the empirically relevant price region, defined as containing all sample observation n = 1,…,N as well as any price points c = 1,…,C that will be used for subsequent analyses and/or simulations based on the estimated model.

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

With respect to Definition 1 it is important to also note that the range of $\boldsymbol{\psi}$ depends on the particular dataset. Moreover, its construction is driven by different estimation purposes:

a. *Interpretation of estimated function as local approximation*: This only requires the smallest possible set $\boldsymbol{\psi}$ at a singular point $\mathbf{p}^0$ in the regressor space. Such a set could be motivated by the construction of the flexible functional form as a series expansion around this singular point and if interest only lies in characterizing such a point.

[8] The requirement that $\boldsymbol{\psi}$ is a closed set simplifies the proofs of some later propositions, but is not necessary for any other reason.

b. *Intended policy simulations*: In the context of our cost function example, $c(\mathbf{p})$, the researcher may be interested in deriving the demand function at point $\mathbf{p}^1$. This point could be just another observed data point from the original data set or a new point of interest for a scenario simulation, for example after adding a tax $\mathbf{t}$ to $\mathbf{p}^0$ (i.e. $\mathbf{p}^1 = \mathbf{p}^0 + \mathbf{t}$) in order to evaluate policy options. In general, the definition of the set $\boldsymbol{\psi}$ defines the suitable (theory consistent) range of regressor values for model simulations.
c. *Technical requirements*: If the estimated function will be used in the context of a larger simulation model, numerical requirements may lead to a rather large set $\boldsymbol{\psi}$ often extending considerably beyond the initial range of interest in economic analysis. For example, to guarantee the proper functioning of iterative algorithms sometimes used in partial equilibrium models, the set $\boldsymbol{\psi}$ has to encompass the range of values occurring in the numerical procedure. An estimated flexible functional form with a large set $\boldsymbol{\psi}$ could also replace globally regular and therefore less flexible functional forms most often employed in Computable General Equilibrium models.
d. *Intended graphical analysis*: For example in order to show a set of empirically estimated isoquants of a production functions, $\boldsymbol{\psi}$ should be at least as large as the envisaged domain of the figure.

2.2 Statistical Model and Bayesian Context

Although the methodology is applicable in other contexts, here we follow the example of the previous section and hence, describe the setting as an estimation of a system of M equations

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

(1) is the empirical specification of 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.[9] We assume that $\boldsymbol{\varepsilon}$ is an $M \cdot N \times 1$ unknown error vector with mean 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[10]

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

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

[10] We use the superscript ‘R’ for a ‘regular’ set, and ‘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 $\boldsymbol{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}_* \cup \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_* = \boldsymbol{\Theta}$.

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

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

where $L(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})$ 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 a priori independent, the joint prior is defined as

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

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 $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ is dependent on the choice of $\boldsymbol{\psi}$, the marginal conditional improper[11] prior on the $\boldsymbol{\beta}$ vector is specified as an indicator function

$$p\left(\boldsymbol{\beta}|\boldsymbol{\psi}\right) = 1\{\boldsymbol{\beta}\in\Theta^{\mathrm{R}}|\boldsymbol{\psi}\} \tag{5}$$

where the prior equals 1 if regularity holds at the value $\boldsymbol{\beta}\,\forall\,\mathbf{p}\in\boldsymbol{\psi}$, 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 $\mathbf{f}(\mathbf{p};\boldsymbol{\beta})$ but also determines the shape of $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ and therefore the potential fit of the economic model to the data. In the remainder of the paper $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ 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, $\mathbf{y} = \mathbf{f}(\mathbf{P};\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$, 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 $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$, 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 $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ on the support $\boldsymbol{\Theta}^{\mathbf{R}}$. The outcomes are then

[11] 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 $p(\boldsymbol{\beta}|\boldsymbol{\psi})$ also includes information about the price space as part of its specification. Also, $\mathbf{1}\{\boldsymbol{\beta}\in\boldsymbol{\Theta}^{\mathbf{R}}|\boldsymbol{\psi}\}$ is technically not a "proper" prior distribution. It is not normalized to integrate to 1, and moreover, if $\boldsymbol{\Theta}^{\mathbf{R}}|\boldsymbol{\psi}$ does not have finite volume, $\int p\left(\boldsymbol{\beta}|\boldsymbol{\psi}\right)\mathrm{d}\boldsymbol{\beta} = \infty$. However our prior effectively indicates the set membership of $\boldsymbol{\beta}$, i.e., if it is regular or not, and it is an uninformative prior on $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$.

used to approximate posterior expectations via the appropriate empirical estimates, e.g. $J^{-1}\sum_{j=1}^{J} \mathrm{g}(\mathbf{b}(\mathrm{j}))$ for approximating $\mathrm{E}[g(\boldsymbol{\beta})]$. The estimates converge to the true expectations as J increases. [12]

To account for the regularity prior $p(\boldsymbol{\beta}|\boldsymbol{\psi})$, the simulator should ensure that any drawn parameter vector $\mathbf{b}^{(j)}$ implies regularity of $\mathbf{f}(\mathbf{p};\boldsymbol{\beta})$ for every point $\mathbf{p}$ in the predefined set $\boldsymbol{\psi}$, i.e. $\mathbf{b}^{(j)} \in \boldsymbol{\Theta}^{\mathbf{R}}|\boldsymbol{\psi}\,\forall j$. Since theoretically there are an infinite number of points in $\boldsymbol{\psi}$, they cannot all be checked explicitly. In general the connectedness can be approximated by a fine grid denoted by the disconnected set $\boldsymbol{\psi}_{\mathrm{g}} \subset \boldsymbol{\psi}$ which consists of possibly $Q \approx$ tens-of-thousands of equidistant distinct points. [13] Within the MCMC an Accept–Reject algorithm is then implemented to guarantee that $\forall\ \mathbf{b}^{(j)}$ the regularity conditions hold for any single of the Q grid point, i.e. that $\mathbf{b}^{(j)} \in \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}} \quad \forall j$, whereby $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}$ is the *approximated* regularity posterior support, which will tend towards the actual set $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ the finer the approximation grid $\boldsymbol{\psi}_{\mathrm{g}}$. In order to circumvent the approximate nature of this representation, in a later subsection (Step 4), we identify problem conditions under which checking certain key points in $\boldsymbol{\psi}$ will guarantee overall regularity $\forall\ \mathbf{p} \in \boldsymbol{\psi}$. Although we cannot always, for all functional forms, universally identify conditions under which $\boldsymbol{\psi}_{\mathrm{g}}$ will automatically guarantee regularity in $\boldsymbol{\psi}$, we view this as in important step in the right direction.

3 Regionally Regular Estimation Procedure

This section describes our proposed method for estimating $f(\mathbf{p};\boldsymbol{\beta})$ subject to the nonlinear inequality constraints $\mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0}\ \forall\ \mathbf{p} \in \boldsymbol{\psi}$. To start we provide a complete stepwise description in Box 1. The procedure consists of three parts: pre-analysis of the problem (Steps 1–4), application of the MHARA (Steps 5–11) and inferences based on the regularity posterior (Step 12). In the subsections to follow, we explain the objectives of the steps[14] and develop necessary technical details.

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 $p(\boldsymbol{\beta},\boldsymbol{\psi}) = 1\{\boldsymbol{\beta} \in \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}\}$: The regularity conditions (defined by economic theory) are identified (Step 2), the

[12] See literature cited in footnote 14 for useful introductions into MCMC methods.

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

[14] *Step 1, Step 5, Step 7, Step 9–11* 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 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 et al. (2000) for a further discussion of appropriate modifications of the Metropolis-Hastings algorithm.

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

Step 1	Estimate $\mathbf{y} = \mathbf{f}(\mathbf{P}; \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$ *without* imposing inequality constraints to obtain the unconstrained estimate $\mathbf{b}_u$ of $\boldsymbol{\beta}$ as well as the estimated $L \times L$ covariance matrix $\mathbf{cov}(\mathbf{b}_u)$.
Step 2	Define $\mathbf{i}()$ that characterizes the regularity conditions for the function being estimated.
Step 3	Define $\boldsymbol{\psi}$according to Definition 1. If the proposed region is not convex, define a sequence of I convex subsets $\boldsymbol{\psi}_i$ such that $\boldsymbol{\psi} = \bigcup_{i=1}^{I} \boldsymbol{\psi}_i$.
Step 4	Selection of evaluation points: For the h^{th} function $i_h(\mathbf{p}; \boldsymbol{\beta})$: analyze which *Properties* I–V hold $\forall\,(\mathbf{p}, \boldsymbol{\beta}) \in (\boldsymbol{\psi} \times \boldsymbol{\Theta})$ and define $\boldsymbol{\psi}_{gh}$ according to Table 1. Repeat Step 4 $\forall$ h.
Step 5	Initialize the Markov Chain with a regular parameter vector: If $\mathbf{b}_\mathrm{u} \in \boldsymbol{\Theta}^\mathrm{R}$, 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 $p(\mathbf{b}^{(*)}; \mathbf{b}^{(j)})$.
Step 7	If $\mathbf{b}^{(*)}$ is irregular at the vertices of $\boldsymbol{\psi}$, set r=0.
Step 8	Repeat Step 4, but instead of evaluating $\mathbf{i}()$ on $(\boldsymbol{\psi} \times \boldsymbol{\Theta})$, evaluate $\mathbf{i}()$ $\forall$ $(\mathbf{p},\mathbf{b}^{(*)}) \in (\boldsymbol{\psi} \times \mathbf{b}^{(*)})$, i.e. conditional on the very last draw $\mathbf{b}^{(*)}$.
Step 9	If $\mathbf{b}^{(*)}$ is regular in $\boldsymbol{\psi}_\mathrm{g}$, calculate $r = p(\mathbf{b}^{(*)}\vert\mathbf{y}, \boldsymbol{\psi})/p(\mathbf{b}^{(j)}\vert\mathbf{y}, \boldsymbol{\psi})$, else set $r = 0$.
Step 10	if $r > 1$, $\mathbf{b}^{(j+1)} = \mathbf{b}^{(*)}$ else if Uniform(0,1) $\leq r$, $\mathbf{b}^{(j+1)} = \mathbf{b}^{(*)}$, 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 $p(\boldsymbol{\beta}\vert\mathbf{y}, \boldsymbol{\psi})$.
Step 12	Analyze $p(\boldsymbol{\beta}\vert\mathbf{y}, \boldsymbol{\psi})$, i.e. calculate point estimates and perform inferences.

empirical relevant region $\boldsymbol{\psi}$ is chosen by the researcher (Step 3) and subsequently approximated by a grid $\boldsymbol{\psi}_\mathrm{g}$ (Step 4).

Step 2: The regularity conditions of $f(\cdot)$ are to be translated into H inequality constraint functions $[i_1, i_2 \ldots, i_H]$ 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 Sect. 2.1.

Step 3: In contrast to defining $\boldsymbol{\psi}$ as one convex hyperrectangle (as in Gallant and Golub 1984; Terrell 1996), we define $\boldsymbol{\psi}$as any connected (possibly non-convex) set. This has potential advantages. First consider the following adaptation of a well-known result from optimization theory:

Lemma 1 *Let $\boldsymbol{\psi}_*$ be any subset of the regressor space $\boldsymbol{\pi}$ and let s: $\boldsymbol{\Theta}^\mathrm{R}\vert\boldsymbol{\psi}_* \rightarrow \Re^1$ be any scalar function.*

$$If\,\boldsymbol{\psi}_{1*} \subset \boldsymbol{\psi}_{2*},\; then \max_{\boldsymbol{\beta}\in\,\boldsymbol{\Theta}^\mathrm{R}\vert\boldsymbol{\psi}_{*1}} s(\boldsymbol{\beta}) \geq \max_{\boldsymbol{\beta}\in\,\boldsymbol{\Theta}^\mathrm{R}\vert\boldsymbol{\psi}_{*2}} s(\boldsymbol{\beta}).$$

Suppose $s\,(\boldsymbol{\beta})$ is any scalar goodness of fit measure maximized when estimating the model. The lemma then states that the fit of a model regular in $\boldsymbol{\psi}_{1*}$ is at least as good as the fit when imposing regularity in $\boldsymbol{\psi}_{2*}$, given that $\boldsymbol{\psi}_{1*} \subset \boldsymbol{\psi}_{2*}$. This suggests that within the context of Definition 1 (see Sect. 2) $\boldsymbol{\psi}$ should be defined as small as possible. This may result in a non-convex set $\boldsymbol{\psi}$. Our methodology can be equally applied by decomposing $\boldsymbol{\psi}$ 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 practice, the construction of $\boldsymbol{\psi}$could be the $I = N + C$ line segments connecting all empirically relevant points, thereby promising an increased fit of the estimated model to the data (in comparison of constructing a convex hyperrectangle around all empirical relevant points). In our application in Sect. 4, we compare these two approaches.

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

Step 4: As outlined in Sect. 2.3, $\boldsymbol{\psi}$ is approximated by $\boldsymbol{\psi}_{\mathrm{g}}$ and regularity is explicitly checked for a high number, 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 problem 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 allows for a reduction of regularity checks to a number $\mathrm{Q}^* < Q$ that

(a) improve the computational speed of the algorithm, while
(b) maintaining the accuracy of the approximation obtained from the original Q-grid.

In order to identify those cases where a reduction in grid points is possible we define the following properties relating to $f(\mathbf{p}; \boldsymbol{\beta})$, $\boldsymbol{\psi}$, and i_h:

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 either $\partial i_{\mathrm{h}} / \partial p_{\mathrm{m}} \geq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$ or $\partial i_{\mathrm{h}} / \partial p_{\mathrm{m}} \leq 0 \; \forall \; \mathbf{p} \in \boldsymbol{\psi}$.*

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

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

If certain combinations of these properties hold, which are summarized as six cases in Table 1, then we perform (depending on Cases 1–6) the regularity checks at one of the following much smaller $Q* < Q$ subsets of $\boldsymbol{\psi}$, which are either, just one vertex, all vertices, the boundary bd($\boldsymbol{\psi}$), one side of, or the 'shield' $\mathbf{S}$*, which are defined as:

15 Subsequently, in order to save notation, the subindex $_i$ is omitted. 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 Sect. 4.2.

Table 1 Sufficient conditions for defining the evaluation set as a subsets of $\boldsymbol{\psi}$

Case	Property I	Property II	Property III	Property IV	Property V	$\boldsymbol{\psi}_h$	Support generated by the hth grid	Proposition
1	+					Boundary $\mathbf{B}_h$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{B}_{gh}\supset\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	1a
2	+	+				One vertex $\mathbf{z}_h$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{z}_h=\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	1b
3			+			Boundary $\mathbf{B}_h$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{B}_{gh}\supset\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	2a
4		+	+			Side $\mathbf{S}_h$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{S}_{gh}\supset\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	2b
5		+		+		All vertices $\mathbf{Z}_h$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{Z}_h=\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	3
6					+	Shield $\mathbf{S}*$	$\boldsymbol{\Theta}^{\mathrm{R}}\vert\mathbf{S}*\supset\boldsymbol{\Theta}^{\mathrm{R}}\vert\boldsymbol{\psi}$	4

In the third to last column the symbol $\boldsymbol{\psi}_h$ is a placeholder for $\mathbf{B}_h, \mathbf{S}_h, \mathbf{S}*, \mathbf{z}_h$, and $\mathbf{Z}_h$. The subindex h indicates that each inequality constraint function i_h requires its own $\boldsymbol{\psi}_h$, all of which are proper subsets of $\boldsymbol{\psi}$

(1) $\mathbf{z}$ is the $K \times 1$ price vector is one vertex of the hyperrectangle $\boldsymbol{\psi}(Q* = 1)$.[16]
(2) $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{2^K}]$ is a $K \times 2^K$ matrix of all vertices of the hyperrectangle $\boldsymbol{\psi}(Q* = 2^K)$.
(3) $\mathbf{B} = \mathrm{bd}\,(\boldsymbol{\psi})$ denotes the boundary of $\boldsymbol{\psi}$.
(4) $\mathbf{S} \subset \mathbf{B}$ is one side of the hyperrectangle orthogonal to the m^{th} price-axis.
(5) $\mathbf{S}* \subset \mathbf{B}$ is a set that can be viewed as a "shield" bounding $\boldsymbol{\psi}$ from below, i.e. from the perspective of rays emanating from 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 $\mathbf{S}* = \{\mathbf{p} \in \mathrm{bd}\,(\boldsymbol{\psi}) : \forall \phi \text{ if } \phi \in \mathrm{bd}\,(\boldsymbol{\psi}) \cap l\,(\mathbf{0}, \mathbf{p})\,, \text{ then } \|\mathbf{p}\| \le \|\phi\|\}$.

With these definitions, the six cases in Table 1 read row-wise as follows:

For Cases 1–5: Suppose for the h^{th} elementary inequality constraint function i_{h} the properties (designated by +) hold:

$$i_h \ge 0 \;\forall\; \mathbf{p} \in \boldsymbol{\psi} \text{ iff } i_h \ge 0 \;\forall\; \mathbf{p} \in \boldsymbol{\psi}_h,$$

whereby $\boldsymbol{\psi}_h$ is a placeholder for one of the evaluation sets, as indicated in the column '$\boldsymbol{\psi}_h$'.

For Case 6: For inequality constraint functions $i^*(\cdot)$ that impose either nonnegative slope, nonpositive slope, concavity and/or convexity:

$$\mathbf{i}^*(\cdot) \ge \mathbf{0}\,\forall\,\mathbf{p} \in \boldsymbol{\psi} \text{ iff } \mathbf{i}^*(\cdot) \ge \mathbf{0}\,\forall\,\mathbf{p} \in \mathbf{S}^*.$$

For the proofs of these statements see Sect. A1 of the appendix.

[16] 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: If $\partial i_h/\partial p_k \le 0 \;\forall\; \mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\max}$ and if $\partial i_h/\partial p_k \ge 0 \;\forall\; \mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\min}$.

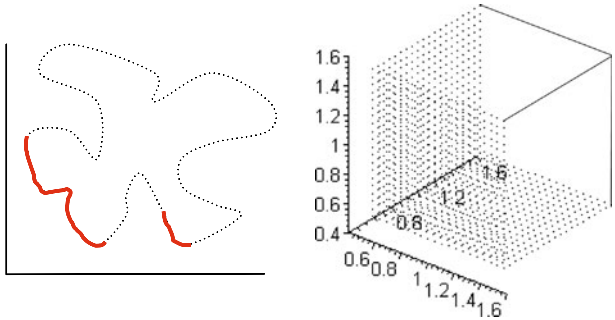


Fig. 2 Illustrations of evaluation grids for the Accept–Reject algorithm. To the left, an example of a shield $\mathbf{S}* \subset \boldsymbol{\psi}$ is displayed. To the right the shield grid $\mathbf{S}_g* \subset \boldsymbol{\psi} = \{\mathbf{p} : \mathbf{p} \in \times_{k=1}^{3}[.5, 1.5]\}$ which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

The Cases 2 and 5 are of special interest because not only these cases enhance the computational speed *considerably* by reducing the evaluation set to include the vertices only, but, and perhaps more importantly, because from a theoretical perspective under Cases 2 and 5, we do not rely anymore on some approximation set $\boldsymbol{\psi}_g$, (that would guarantee regularity only if $Q \to \infty$)[17], but it is guaranteed that regularity of f holds also between all of the grid points. This is expressed in the following proposition:

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

For all other cases, clearly, 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 '$_g$').

Finally note, that the first five cases in Table 1 are very general in the sense that any kind of shape restriction could be imposed. Case 6, instead, is less general, but applies to many functions in economics as these are often homothetic and constrained by monotonocity and or curvature conditions (this suits our cost-function example in Sects. 2 and 4).

3.2 The Metropolis-Hastings Accept–Reject Algorithm

Steps 6–11 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 $j^{\text{th}} + 1$ vector in the MCMC sequence $\{\mathbf{b}^{(j)}\}_{j=1}^{J+S}$, is generated by a symmetric *proposal distribution* $p(\mathbf{b}^{(*)}; \mathbf{b}^{(j)})$. We use the multivariate normal distribution $\mathbf{N}(\mathbf{b}^{(j)}, \delta\mathbf{cov}(\mathbf{b}_{\text{u}}))$ to first generate the $L \times 1$ vector $\mathbf{b}^{(**)}$

[17] If $Q \to \infty$, i.e. the number of equidistant grid points of $\boldsymbol{\psi}_g$ goes to infinity, and $\mathbf{i}(\cdot)$ is continuously differentiable, then any parameter value $\mathbf{b} \in \boldsymbol{\Theta}^{\text{R}}|\boldsymbol{\psi}_g$ is such that $f(\mathbf{p}; \mathbf{b})$ is almost everywhere in $\boldsymbol{\psi}$ regularity-retaining.

whereby the covariance matrix $\mathbf{cov}(\mathbf{b}_u)$ is scaled by the scalar δ such that approximately 25% of the draws get accepted in Step 10.[18] Since $\mathbf{b}^{(**)}$doesn't account for the *linear equality* constraints on the parameters (e.g. the symmetry condition on the Hessian $\partial^2 f(\mathbf{p};\boldsymbol{\beta})/\partial\mathbf{p}\partial\mathbf{p}\prime$), we further calculate the linearly restricted draw of the *L-V free* parameters $\mathbf{b}^{(*)}$ as

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

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 $\mathbf{b}^{(*)}$. Note that now $\mathbf{b}^{(*)}$ accounts for the *linear equality* constraints, but not yet for the nonlinear *inequality* constraints $\mathbf{i}(\mathbf{p};\boldsymbol{\beta})\geq\mathbf{0}$, which is the task of the next step.

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 in Step 4 in some h^{th} evaluation the evalution set equals $\mathbf{Z}_h$ or or $\mathbf{z}_h$, 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 1–11 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 $E[\boldsymbol{\beta}]$ of the regularity posterior.[19] However, this may result in regularity violations, as indicated in the following proposition.

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

Although the mean point estimator $E[\boldsymbol{\beta}]$ has some desirable properties under squared error loss, if it is found that $E[\boldsymbol{\beta}]$ is irregular, this undermines the estimation objective of imposing the regularity constraints. Hence, here we propose two alternative estimators that, in addition to imposing regularity (objective I), maximize a model fit measure $s(\boldsymbol{\beta})$ on $\boldsymbol{\Theta}^R|\boldsymbol{\psi}_g$, 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.

[18] The term *proposal distribution* stems from the fact that $p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$ *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 $\mathbf{b}^{(j)}$. See Robert and Casella (2002, pp. 281–283) discussing the choice of δ.

[19] These include Barnett et al. (1991), Koop et al. (1994), Koop et al. (1997), Terrell (1996), Terrell and Dashti (1997), O'Donnell et al. (1999), Griffiths et al. (2000), Chua et al. (2001), Kleit and Terrell (2001), Cuesta et al. (2001), Adkins et al. (2002), O'Donnell et al. (2001), and O'Donnell and Coelli (2003).

Under the assumption of a normal error distribution, we present the mode

$$\boldsymbol{\beta}^{(\mathrm{mode})} = \underset{\boldsymbol{\beta}\in\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}}{\arg\max}\left\{p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}_{\mathrm{g}})\right\}.$$

of the regularity posterior as the point estimate which maximizes the model fit subject to the regularity conditions. To motivate $\boldsymbol{\beta}^{(mode)}$, note that the information contained in the normal unrestricted posterior $p_{\mathrm{u}}(\boldsymbol{\beta}|\mathbf{y}) \propto |(N-L)\,\boldsymbol{\Sigma}|^{-N/2}$ (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) proves that the minimization of $|\boldsymbol{\Sigma}|$ is equivalent to Maximum Likelihood (ML) estimation in the case of the nonlinear normal classical SUR model. Since $(N-L)$ and the exponent $-N/2$ are fixed constants, the minimization of $|\boldsymbol{\Sigma}|$ over $\boldsymbol{\beta}\in\boldsymbol{\Theta}$ produces the exact same result as the maximization of $p_u(\boldsymbol{\beta}|\mathbf{y})$ over $\boldsymbol{\beta}\in\boldsymbol{\Theta}$. So long as no other prior than the regularity prior is applied, we have that $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto p_{\mathrm{u}}(\boldsymbol{\beta}|\mathbf{y})\cdot 1\{\boldsymbol{\beta}\in\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}\} \propto |(N-L)\,\boldsymbol{\Sigma}|^{-N/2}$. Thus the normal classical inequality-constrained-ML estimator generates a point estimate that is numerically equivalent to the mode of $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$. In order to approximate the solution based upon the MCMC outcomes $\{\mathbf{b}^{(j)}\}_{j=S+1}^{J+S}$, one can simply compare the values $p_{\mathrm{u}}(\mathbf{b}^{(j)}|\mathbf{y})\;\forall\, j$ resulting from the MHARA as

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

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

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

which minimizes the posterior weighted deviation over $\boldsymbol{\beta}\in\boldsymbol{\Theta}^{\mathrm{R}}$, where $||\cdot||_{\varphi}$ is some vector norm[20] measuring the distance between two points within $\boldsymbol{\Theta}^{\mathrm{R}}$. For example, with $||\cdot||_2$, the standard Euclidean norm, $\mathbf{b}^{(\mathrm{LF}_2)} = \underset{\mathbf{b}^{(j)}}{\arg\min}\, J^{-1}\sum_{i=1}^{J}(\mathbf{b}^{(j)}-\mathbf{b}^{(i)})'(\mathbf{b}^{(j)}-\mathbf{b}^{(i)})$, minimizing 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 $\mathbf{b}^{(mode)}$ are members of the regular set $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ 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 $\boldsymbol{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}$.

[20] Given an N-dimensional $\mathbf{x}$ a general vector norm $||\mathbf{x}||_{\varphi}$, for $\phi = 1, 2, \ldots$ is a nonnegative defined as $||\mathbf{x}||_{\varphi} = \left[\Sigma_{n=1}^{N}|\mathbf{x}|^{\varphi}\right]^{1/\varphi}$. 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 $||\mathbf{x}||_2 = [\Sigma_{n=1}^{N}\mathbf{x}^2]^{1/2}$.

The proposed methodology is general enough to be adopted in both the Bayesian and the Classical framework. In the Classical framework, we could maximize a likelihood function subject to (non-)linear inequality constraints $\mathbf{i}()$ and the point estimate is the mode of the MCMC-simulated likelihood, which generally will be identical to the mode, $\boldsymbol{\beta}^{(\text{mode})}$, of the regularity posterior. The suggested LF criterion, leading to $\boldsymbol{\beta}^{(\text{LF}_\varphi)}$, 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).[21] In the first subsection local, global and regional regularity approaches are compared based on a specified convex set $\boldsymbol{\psi}^{\square}$. (We use the superscript symbol $^{\square}$ to denote a convex set.) The purpose of the second subsection is to demonstrate the effects of shrinking the size of $\boldsymbol{\psi}$.

4.1 Experiment I—Convex Cube $\boldsymbol{\psi}$

4.1.1 Data Generation

We now briefly describe the design of the simulations.[22] The *true* data generation process is formulated by the well-known CES cost function $f^{\text{CES}}(\mathbf{p}; \alpha_k, \rho) = \left[\Sigma_{k=1}^{3} a_k^{1/(1-\rho)} \cdot p_k^{-\rho/(1-\rho)}\right]^{(1-\rho)/-\rho}$. As in Terrell, no stochastic error term is added. The derivatives result, by Shephard's Lemma, in $K = 3$ input demand functions,

$$x_k = \partial f^{\text{CES}}/\partial p_k = [\alpha_k \cdot f^{\text{CES}}/p_k]^{1/(1-\rho)} \tag{5}$$

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 (6) this produces $64 \cdot 3$ true input demand levels, where $\mathbf{x}_k$ is 64×1 with $k = 1, 2, 3$.

[21] The model is kept rather basic 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 $\mathbf{p}$.

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

Table 2 Global, regional and local approach—comparison based on AIM cost functions

Model	Forecast error and regularity violations evaluated over $\boldsymbol{\psi}_g^{\square}$	Estimation approach			
		Local regularity[a]	Global regularity[a]	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%

Experiment based on Tables 1 and 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 $8000^{-1}\Sigma_{g=1}^{8000} x_{gk} = 0.2552\,\forall\, k$ and standard deviation of $\text{std}(\mathbf{x}_k) = 0.2230\,\forall\, k$ over the evaluation grid $\boldsymbol{\psi}_g^{\square}$

[a] Some considerable differences exist between our and Terrell (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

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, $\hat{x}_k = \partial \mathbf{f}_k^{\text{AIM}(\tau)}\left(\mathbf{P}, \hat{\boldsymbol{\beta}}\right)/\partial p_{k+\hat{\mathbf{u}}_k}$ is estimated, whereby $\hat{\text{u}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ represents the 64×1 approximation error vector to the 'true' data generation process (7), $L < N$ [23] and $\hat{\mathbf{x}}_k$ is the estimated $k^{\text{th}} 64 \times 1$ input demand vector derived from the Asymptotically Ideal Production Model, AIM(τ), with

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

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

which are homogenous of degree one, constant returns to scale unit cost functions.[24]

As in Terrell (1995), the performance of the AIM(τ) is evaluated over the cubic region $\boldsymbol{\psi}^{\square} = \{\mathbf{p} : \mathbf{p} \in \quad \boldsymbol{\psi}^{\square} = \left\{\mathbf{p} : \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\right\}$ by defining a grid $\boldsymbol{\psi}_{\mathrm{g}}^{\square} \subset \boldsymbol{\psi}^{\square}$ of 20 equidistant prices for each input. Thus in total $\boldsymbol{\psi}_{\mathrm{g}}^{\square}$ 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*, $\mathrm{MAE}_k = \mathrm{MAE}_k = sgn \left\{ \hat{u}_{\arg\max\limits_q \{|\hat{u}_{qk}|\},k} \right\} \cdot \max\limits_q \{ \left| \hat{u}_{qk} \right| \}$, and (b) the *average absolute approximation error*, $\mathrm{AAAE}_k = Q - ^{1} \sum_{q=1}^{Q} |\hat{u}_{qk}|$, 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 optimizing the model fit MAE and AAAE values close to zero are preferred.

4.1.3 Results

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 (1995, Tables 1 and 2, pp. 9–10) simulation experiment, and the last two columns apply the method described in Sect. 4.

First the demand system is estimated subject to local concavity and monotonicity constraints guaranteeing regularity for the underlying AIM(τ) cost function at $\mathbf{p}^{\mathrm{M}} = [1, 1, 1]$, 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 $\mathbf{p}^{\mathrm{M}}$ (leading to regularity violations of about 20% of the grid points), which is illustrated in Fig. 3. It is particularly instructive to note that the monotonicity violations are substantially more frequent than the concavity violations, which is disconcerting given that Terrell, and in fact most researchers in similar previous studies, did not check for monotonicity violations (see Barnett 2002).

In the column 'global regularity' economic theory holds globally on $\boldsymbol{\pi}$ through the imposition of nonnegativity constraints on all the AIM parameters $\boldsymbol{\beta}$ (as in Terrell 1995) which confirms numerically the result of Lemma 1 by showing a decreased model fit.

The last two columns show the MHARA[25] results imposing the regularity conditions regionally on $\boldsymbol{\psi}^{\square}$. First we take the mean—as is commonly done—as the point estimate for $\boldsymbol{\beta}$. As one might expect this 'regional mean approach' leads to improved

[24] 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, $\tau \rightarrow \infty$, 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 et al. (1991).

[25] For MCMC sampling in the context of the normal SUR model, we want to refer to the very useful exposition by Griffiths 2003.

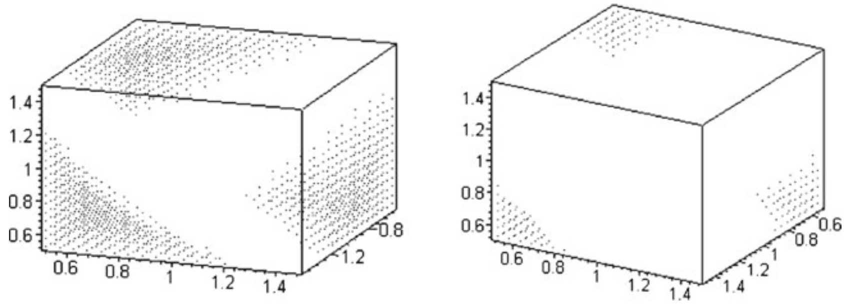


Fig. 3 Violations on the price grid $\boldsymbol{\psi}_g^{\square}$ *in the case of the local regularity approach.* . In 19.09% of the grid points monotonicity is violated (left cube) and in 3.11% concavity is violated (right cube). Each black dot is one grid point where violation occurs

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 $\boldsymbol{\beta}$, 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 Sect. 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–V, the maximum of 1142 grid points of the set $\mathbf{S}_g^* \subset \boldsymbol{\psi}_g^{\square}$ had to be evaluated only. Furthermore, for the AIM(1) often only one vertex had to be assessed. This significantly decreased the computational burden compared to previous approaches.

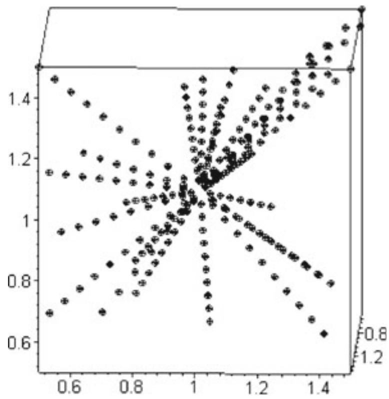
Summarizing Table 2, imposing local regularity increases the model fit in all specifications at the cost of violating monotonicity and concavity within $\boldsymbol{\psi}$, which produces estimation results that are problematic in terms of economic interpretation and further analysis. Imposing regional 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.

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

Fig. 4 The String grid $\boldsymbol{\psi}_g^{\text{string}}$



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 $\boldsymbol{\psi}^{\square}$.[26] 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 $\boldsymbol{\psi}_g^{\square}$ and

(b) $\boldsymbol{\psi}^{\text{string}} = \bigcup_{i=1}^{29} \boldsymbol{\psi}_i$, which covers all $30 = I + 1$ price points by connecting 29 straight lines $\boldsymbol{\psi}_i, i = 1, \ldots I$, between $\mathbf{p}^{\text{M}}$ (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 $\mathbf{p}^{\text{M}}$ 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 $\boldsymbol{\psi}_g^{\square}$.

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 $\boldsymbol{\psi}_g^{\square}$ versus imposing the regularity conditions on $\boldsymbol{\psi}_g^{string}$ 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.

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

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

Model	Model performance statistics		Estimation approach						imposed on $\psi_g^{\square}$ (cube approach)			imposed on ψ_g^{string} (string approach)			Percentage change in error statistics of the string approach relative to the cube 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_g^N	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_g^C	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	$\psi_g^{\square}$	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%				

Table 3 continued

Model	Model performance statistics		Estimation approach						imposed on $\psi_g^{\square}$ (cube approach)			imposed on ψ_g^{string} (string approach)			Percentage change in error statistics of the string approach relative to the cube 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(2)	AAAE	ψ_g^N	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_g^C	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	$\psi_g^{\square}$	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 $a_i = 1$; $\rho = 0.75$

We also supply performance statistics for the string approach evaluated over the cube grid $\psi_g^{\square}$. We do not necessarily advocate such an approach (i.e. defining ψ on a subset of the region where subsequent inferences will be drawn). We rather include these results[27] to again emphasize the trade off between flexibility and regularity: The regional regularity approach can become useless when ψ does not cover the empirically relevant region (because it is likely that outside of ψ 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

This paper develops a procedure for estimating parametric functions subject to regularity conditions derived from economic theory that are imposed on a regular region of the function's domain defined by the analyst. Our method extends upon Terrell (1996) work leading to improved model fit, and is also computationally much faster and more efficient than previous approaches while imposing both curvature and monotonicity on the entire selected region of the regressor space. In fact the generality of the method makes it applicable as a new procedure for the broader problem of estimating regression functions subject to nonlinear inequality constraints.

Our numerical examples illustrate that the tractability of the estimation procedure is enhanced through a reduction in the number of regularity checks required in the estimation process. Another objective was to improve in- and out-of-sample forecasts. The theoretical and numerical results provide evidence that the model fit statistics significantly improve by a) using the posterior mode of the parameters and/or by b) allowing the desired regular region, ψ, to be some connected non-convex set. Finally we demonstrated that the commonly used posterior mean may be inappropriate as a point estimate. For both of the latter problems we suggested simple consistent alternatives.

It will be interesting to compare these results with the currently developing new techniques in nonparametric estimation that imposes shape restrictions. This is to be explored in future research. We hope that the methods and results demonstrated in this paper promote tractability and facilitate efficiency in the analysis of regularity-preserving economic models.

Acknowledgement 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 German Academic Exchange Service (DAAD) for financial support.

[27] 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_g^N, this is not necessarily true when evaluated over the cubic region $\psi_g^{\square}$, (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 $\psi^{\square} \setminus \psi^{\text{string}}$ are calculated by (7).

Appendix

The appendix is divided into three parts. Part A) contains the proofs of the propositions outlined in Table 1 and some further explanations. Part B) lists the remaining proofs of Lemma 1, Propositions 5 and 6 and Part C) provides the data used in Sect. 4.2.

Part (A): Proof of Propositions Outlined in Table 1 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 $\mathbf{b}^{(*)} \in \boldsymbol{\Theta}$, the orthant of strictly positive prices $\boldsymbol{\pi}$ can always be partitioned into two disjoint subsets, $\boldsymbol{\pi}^{\mathrm{R}}|\mathbf{b}^{(*)} \cup \boldsymbol{\pi}^{I\mathrm{R}}|\mathbf{b}^{(*)} = \boldsymbol{\pi}$. We say that $f(\mathbf{p}; \mathbf{b}^{(*)})$ is well behaved on the regular price set $\boldsymbol{\pi}^{\mathrm{R}}|\mathbf{b}^{(*)} = \{\mathbf{p} : \mathbf{i}(\mathbf{p}; \mathbf{b}^{(*)}) \geq \mathbf{0}\,\forall\,\mathbf{p} \in \boldsymbol{\pi}\}$. (2) Since we are particularly interested in the behavior of the function within the set $\boldsymbol{\psi}$, let us define $\boldsymbol{\psi}^{\mathrm{R}} = \boldsymbol{\psi}^{\mathrm{R}}|\mathbf{b}^{(*)} = \{\mathbf{p} : \mathbf{i}(\mathbf{p}; \mathbf{b}^{(*)}) \geq \mathbf{0}\,\forall\,\mathbf{p} \in \boldsymbol{\psi}\} \subset \boldsymbol{\pi}^{\mathrm{R}}|\mathbf{b}^{(*)}$. It has the following features: If $f(\mathbf{p}; \mathbf{b}^{(*)})$ is regular $\forall\,\mathbf{p} \in \boldsymbol{\psi}$, then $\boldsymbol{\psi}^{\mathrm{R}} = \boldsymbol{\psi}$. In general, however, $\boldsymbol{\psi}^{\mathrm{R}} \subset \boldsymbol{\psi}$. For Propositions 1a to 2b and 4, we prove sufficiency by contrapositive. To prove necessity is trivial and is omitted.

Proposition 1a

$$\textit{Suppose}\left\{\begin{array}{lll} \partial i_{\mathrm{h}}/\partial p_1 \geq 0\,\forall & \mathbf{p} \in \boldsymbol{\psi} & \{or\ \partial i_{\mathrm{h}}/\partial p_1 \leq 0\,\forall\,\mathbf{p} \in \boldsymbol{\psi}\} \\ \partial i_{\mathrm{h}}/\partial p_2 \geq 0\,\forall & \mathbf{p} \in \boldsymbol{\psi} & \{or\ \partial i_{\mathrm{h}}/\partial p_2 \leq 0\,\forall\,\mathbf{p} \in \boldsymbol{\psi}\} \\ : & : & : \\ : & : & : \\ \partial i_{\mathrm{h}}/\partial p_{\mathrm{K}} \geq 0\,\forall & \mathbf{p} \in \boldsymbol{\psi} & \{or\ \partial i_{\mathrm{h}}/\partial p_2 \leq 0\,\forall\,\mathbf{p} \in \boldsymbol{\psi}\} \end{array}\right\} (\textit{Property}\ \mathrm{I\ holds})$$

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

Proof of Proposition 1a [28]: Suppose not, then $\exists\ \mathbf{p}* \in \boldsymbol{\psi}^{\mathrm{IR}}\backslash\mathbf{B}$ with $i_h\,(\mathbf{p}*) < 0$. Further $\exists\ \mathbf{p}^{\mathrm{B}} = \left[p_1^{\mathrm{B}}, p_2^{\mathrm{B}}, \ldots, p_K^{\mathrm{B}}\right]^T \in \mathbf{B}$ which has the following property:

$$\begin{array}{ll} p_1^{\mathrm{B}} \leq p_1^* & \{\text{or } p_1^{\mathrm{B}} \geq p_1^*\} \\ p_2^{\mathrm{B}} \leq p_2^* & \{\text{or } p_2^{\mathrm{B}} \geq p_2^*\} \\ : & : \\ p_K^{\mathrm{B}} \leq p_K^* & \{\text{or } p_K^{\mathrm{B}} \geq p_K^*\} \end{array}$$

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

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

We conclude that only $\mathbf{B} \subset \boldsymbol{\psi}$ has to be evaluated if *Property* I holds. In practice, however, we cannot check for the connected set but approximate it by $\mathbf{B}_{\mathrm{g}}$, thus still running the risk of violating regularity in the neighborhood of the points in $\mathbf{B}_{\mathrm{g}}$. Fortunately however, in many applications we can apply the results of the following proposition.

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

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

$$\begin{array}{ll} z_1 \leq p_1^{\mathrm{B}} & \{\text{or } z_1 \geq p_1^{\mathrm{B}}\} \\ z_2 \leq p_2^{\mathrm{B}} & \{\text{or } z_2 \geq p_2^{\mathrm{B}}\} \\ \vdots & \quad\vdots \\ z_K \leq p_K^{\mathrm{B}} & \{\text{or } z_K^{\geq} p_K^{\mathrm{B}}\} \end{array}$$

Hence $i_h(\mathbf{z}) \leq i_h(\mathbf{p}^{\mathrm{B}}) \leq i_{\mathrm{h}}(\mathbf{p}*) < 0$. So $\mathbf{z} \in \boldsymbol{\psi}_h^{\mathrm{IR}}$. □

Since—under the conditions *Properties* I and II—whenever $[p_1^{\min\{\max\}}, p_2^{\min\{\max\}}, \ldots, p_K^{\min\{\max\}}]^{\mathrm{T}} \in \boldsymbol{\psi}_h^{\mathrm{R}}$, then $\boldsymbol{\psi}_h^{\mathrm{R}} = \boldsymbol{\psi}$, we conclude that only this single vertex point has to be checked.[29] If for some inequality constraint function i_{h} *Property* I does not hold, but instead the relaxed version Property III, then the following result still greatly simplifies the Accept–Reject algorithm.

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

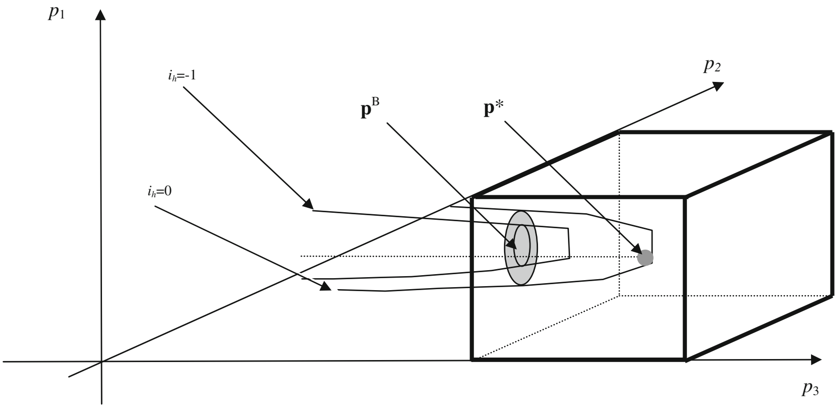
Proof of Proposition 2a Suppose not, then $\exists\, \mathbf{p}^* \in \boldsymbol{\psi}^{\mathrm{IR}} \backslash \{\mathbf{B}\}$ with $i_h(\mathbf{p}*) < 0$. Further $\exists\, \mathbf{p}^{\mathrm{B}} = [p_1^{\mathrm{B}}, p_2^{\mathrm{B}}, \ldots, p_K^{\mathrm{B}}]^T \in \mathbf{B}$ which has the following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

$$\begin{aligned} p_m^{\mathrm{B}} &\leq p_m^* \{\text{or} \, p_m^{\mathrm{B}} \geq p_m^*\} \\ \mathbf{p}_{-m}^{\mathrm{B}} &= \mathbf{p}_{-m}^* \end{aligned}$$

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 $\mathbf{B} \not\subset \boldsymbol{\psi}_h^{\mathrm{R}}$. □

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 inequality constraint functions, in which case we can omit checking the interior of $\boldsymbol{\psi}$. Similarly to Proposition 1b, the following will further enhance the speed of MHARA.

[29] 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.



A 1 Inequality Constraint Function Level Sets $i_h = -1$ and $i_h = 0$ in price space $\boldsymbol{\pi}$. If *properties* II and III hold, $\mathbf{p}^*$ is irregular, and $\partial i_h/\partial p_3 \geq 0$, then the boundary side $\mathbf{S}$ facing towards the p_1–p_2 level contains irregular points $\mathbf{p}^{\mathrm{B}} \in \mathbf{S}^{\mathrm{IR}} \subset \mathbf{S}$ as well. $\mathbf{S}^{\mathrm{IR}}$ is shaded in grey. The set $\boldsymbol{\psi} \subset \boldsymbol{\pi}$ is indicated by the cube

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 $p_m^S = p_m^{\min\{\max\}} \quad \forall\,(p_m^{\mathrm{S}}, \mathbf{p}_{-m}^{\mathrm{S}}) \in \mathbf{S}$. Suppose Properties II and 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. □

In other words, if *Properties* II and III hold, then it is only necessary to evaluate $\mathbf{S}$ which is the side of the hyperrectangle 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.

The following proposition provides sufficiency conditions to check only the extreme points $\mathbf{Z}_h^{\mathrm{e}}$ of a convex set $\boldsymbol{\psi}$. [30] The result does not rely on *Property* II and is hence more general than Case 5 of Table 1. If $\boldsymbol{\psi}$ is a hypercube, then $\mathbf{Z}_h^{\mathrm{e}}$ is equivalent to the 2^K vertices defined in Sect. 3.1 as $\mathbf{Z}_h$.[31]

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

Proof of Proposition 3 A quasi-concave function i_h has the property that its upper contour set $\mathbf{U}_\omega = \{\mathbf{p} : i_h \geq \omega, \mathbf{p} \in \boldsymbol{\psi}, \omega \in \Re^1\}$ is convex. $\boldsymbol{\psi}_h^{\mathrm{R}} = \{\mathbf{p} : i_h \geq 0, \mathbf{p} \in \boldsymbol{\psi}\}$ is an upper contour set $\mathbf{U}_0$ evaluated at $\omega = i_h = 0$ such that $\mathbf{Z}_h^{\mathrm{e}} \in \boldsymbol{\psi}_h^{\mathrm{R}}$ (by assumption). Since, by *Property* IV, $\boldsymbol{\psi}$is convex it follows that $\boldsymbol{\psi}_h^{\mathrm{R}} = \mathbf{U}_0 \cap \boldsymbol{\psi}$ is convex (since the intersection of convex sets is convex). Finally, since any convex set is connected and $\mathbf{Z}_h^{\mathrm{e}} \in \boldsymbol{\psi}_h^{\mathrm{R}}$, it follows that $\boldsymbol{\psi}_h^{\mathrm{R}} = \boldsymbol{\psi}$. □

---

[30] $\mathbf{z}^{\mathrm{e}}$ is an extreme point of $\boldsymbol{\psi}$ iff $\mathbf{z}^{\mathrm{e}} = \lambda \cdot \mathrm{p}_1 + (1-\lambda)\,\mathrm{p}_2,\ \forall\ \mathrm{p}_1, \mathrm{p}_2 \in \boldsymbol{\psi}, \lambda \in (0, 1)$, implies $\mathbf{z}^{\mathrm{e}} = \mathrm{p}_1 = \mathrm{p}_2$.

[31] 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.

Remark 1 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, 1994, pp. 523–531).

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

Proof of Proposition 4 Suppose not, then $\exists\ \mathbf{p}^* \in \boldsymbol{\psi}^{\mathrm{IR}} \backslash \mathbf{S}*$ for which either (a) nonpositive 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, $\alpha \in \Re^1$, 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 relative 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 $\mathbf{S}^*$.

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

Part (B): Proof of Lemma 1, Propositions 5 and 6

Proof of Lemma 1 The proof follows immediately from the definition of $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}* = \{\boldsymbol{\beta} : i(p;\boldsymbol{\beta}) \geq 0\, \forall \mathbf{p} \in \boldsymbol{\psi}*, \boldsymbol{\beta} \in \boldsymbol{\Theta}\}$ which implies that ceteris paribus, the larger the constraining set $\boldsymbol{\psi}* \subset \boldsymbol{\pi}$, the smaller is the support $\boldsymbol{\Theta}^{\mathrm{R}}$, i.e. if $\boldsymbol{\psi}_1* \subset \boldsymbol{\psi}_2*$, then $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_1* \supset \quad \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_2*$. Consequently, maximizing s($\boldsymbol{\beta}$) over the smaller set $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_2*$ can only lead to objective values equal or smaller than as maximizing s($\boldsymbol{\beta}$) over $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_1^*$.
□

Proof of Proposition 5 The proof follows directly from the Propositions 1b and 3 and noting that if the evaluation sets are finite, the regularity posterior can be simulated with support $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi} = \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}$, 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. □

Proof of Proposition 6 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 $\boldsymbol{\Theta}^{\mathrm{R}}$. □

Part (C)

Table 4 Input price observations and out of sample points used for experiment II

n	Input price 1	Input price 2	Input price 3
26×3 input price observation matrix **P**			
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	Input price 1	Input price 2	Input price 3
$C = 4$ scenario input price vectors			
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, L. C., Rickman, D. S., & Hameed, A. (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.

Aït-Sahalia, Y., & Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics, 116*, 9–47.
Barnett, W. A. (1976). Maximum likelihood and iterated Aitken estimation of non-linear systems of equations. *Journal of the American Statistical Association, 71*, 354–360.
Barnett, W. A. (1985). The Minflex-Laurent translog flexible functional form. *Journal of Econometrics, 30*, 33–44.
Barnett, W. A. (2002). Tastes and technology: Curvature is not sufficient for regularity. *Journal of Econometrics, 108*, 199–202.
Barnett, W. A., Geweke, J., & Wolfe, M. (1991). Seminonparametric bayesian estimation of the asymptotically ideal production model. *Journal of Econometrics, 49*(1), 5–50.
Barnett, W. A., Kirova, M., & Pasupathy, M. (1995). Estimating policy-invariant deep parameters in the financial sector when risk and growth matter. *Journal of Money, Credit Bank, 27*(4), part 2, 1402–1430.
Barnett, W. A., & Pasupathy, M. (2003). Regularity of the generalized quadratic production model: A counterexample. *Econometric Reviews, 22*(2), 135–154.
Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo methods in bayesian computation*. New York: Springer.
Chib, S., & Greenberg, E. (1996). Markov Chain Monte Carlo methods in econometrics. *Econometric Theory, 12*, 409–431.
Chua, C. L., Griffiths, W. E., & O'Donnell, C. J. (2001). Bayesian model averaging in consumer demand systems with inequality constraints. *Canadian Journal of Agricultural Economics, 49*, 269–291.
Cuesta, R. A., O'Donnell, C. J., Coelli, T. J., & Singh, S. (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. Armidale: School of Economics, University of New England.
Diewert, W. E., & Wales, T. J. (1987). Flexible functional forms and global curvature conditions. *Econometrica, 55*, 43–68.
Diewert, W. E., & Wales, T. J. (1991). *Multiproduct cost functions and subadditivity tests: A critique of the Evans and Heckman research on the U.S. Bell Systems*. Discussions Paper 91-21 at the Department of Economics, University of British Columbia, Vancouver. http://faculty.arts.ubc.ca/ediewert/9121.pdf.
Evans, D. S., & Heckman, J. J. (1984). A Test for Subadditivity of the Cost Function with an Application to the Bell System. *American Economic Review, 74*, 615–623.
Fischer, D., Fleissig, A. R., & Serletis, A. (2001). An empirical comparison of flexible demand system functional forms. *Journal of Applied Econometrics, 16*(1), 59–80.
Fleissig, A. R., Kastens, T., & Terrell, D. (1997). Semi-nonparametric estimates of substitution elasticities. *Economic Letters, 54*, 209–215.
Fleissig, A. R., Kastens, T., & Terrell, D. (2000). Evaluating the semi-nonparametric Fourier, AIM, and neural networks cost functions. *Economic Letters, 68*, 235–244.
Gallant, A. R., & Golub, G. H. (1984). Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics, 26*, 295–322.
Griffiths, W. E. (2003). Bayesian inference in the seemingly unrelated regressions model. In D. E. A. Giles (Ed.), *Computer-aided econometrics* (pp. 263–290). New York: Marcel Dekker.
Griffiths, W. E., O'Donnell, C. J., & Tan-Cruz, A. (2000). Imposing regularity conditions on a system of cost and factor share equations. *Australian Journal of Agricultural and Resource Economics, 44*, 107–127.
Griffiths, W. E., Skeels, C. L., & Chotikapanich, D. (2002). Sample size requirements for estimation in SUR Models. In A. Ullah, A. Chaturvedi, & A. Wan (Eds.), *Handbook of Applied Econometrics and Statistical Inference.* (pp. 575–590). New York: Marcel Dekker.
Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of the American Statistical Association, 49*, 598–619.
Ivaldi, M., Ladoux, N., Ossard, H., & Simioni, M. (1996). Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers. *Journal of Applied Econometrics, 11*(6), 649–667.
Kleit, A. N., & Terrell, D. (2001). Measuring potential efficiency gains from deregulation of electricity generation: A bayesian approach. *The Review of Economics and Statistics, 83*(3), 523–530.
Koop, G., Osiewalski, J., & Steel, M. F. J. (1994). Bayesian efficiency analysis with a flexible form: The AIM cost function. *Journal of Business and Economic Statistics, 12*(3), 339–346.

Koop, G., Osiewalski, J., & Steel, M. F. J. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics, 76*, 77–105.
Lau, L. J. (1978). Testing and imposing monotonicity, convexity, and quasi-convexity constraints. In M. Fuss & D. McFadden (Eds.), *Production economics: A dual approach to theory and applications (volume 1)* (pp. 409–453). Amsterdam: North-Holland.
Lau, L. J. (1986). Functional forms in econometric model building. In Z. Griliches & M. D. Intriligator (Eds.), *Chapter 26 in handbook of econometrics* (Vol. 3, pp. 1515–1566). Amsterdam, North Holland: Elsevier Science.
Magnus, J. R. (1979). Substitution between energy and non-energy inputs in the Netherlands 1950–1976. *International Economics Review, 20*, 465–484.
Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. USA: Oxford University Press.
Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics* (Vol. 4, Chap. 42, pp. 2524–2558). Amsterdam: North-Holland Pub Co.
O'Donnell, C. J., & Coelli, T. (2003). *A bayesian approach to imposing curvature on distance functions*. Paper presented at the Australasian Meeting of the Econometric Society, Sydney 2003.
O'Donnell, C. J., Rambaldi, A. N., & Doran, H. E. (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., Shumway, C. R., & Ball, V. E. (1999). Input demands and inefficiency in U.S. agriculture. *American Journal of Agricultural Economics, 81*, 865–880.
Racine, J. S., & Parmeter, C. F. (2008). Constrained nonparametric kernel regression estimation and inference. Working paper available at http://www1.maxwell.syr.edu/uploadedFiles/econ/kernel_cons.pdf?n=6322.
Ryan, D. L., & Wales, T. J. (1998). A simple method for imposing local curvature in some flexible consumer demand systems. *Journal of Business and Economic Statistics, 16*(3), 331–338.
Salvanes, K. G., & Tjøtta, S. (1998). A Note on the Importance of Testing for Regularities for Estimated Flexible Functional Forms. *Journal of Productivity Analysis, 9*, 133–143.
Simon, C. P., & Blume, L. (1994). *Mathematics for economists*. New York: W.W. Norton.
Terrell, D. (1995). Flexibility and regularity properties of the asymptotic ideal production model. *Econometric Reviews, 14*(1), 1–17.
Terrell, D. (1996). Incorporating monotonicity and concavity conditions in flexible functional forms. *Journal of Applied Econometrics, 11*, 179–194.
Terrell, D., & Dashti, I. (1997). *Imposing monotonocity and concavity restrictions on stochastic frontiers*. Working Paper, Department of Economics, Louisiana State University, E. J. Ourso College of Business Administration.
Tripathi, G. (2000). Local semiparametric efficiency bounds under shape restrictions. *Econometric Theory, 16*, 729–739.
Zellner, A. (1971). *An introduction to bayesian inference in econometrics*. New York.: John Wiley and Sons.