

CAN WE CLOSE THE GAP BETWEEN THE EMPIRICAL MODEL AND ECONOMIC THEORY?

AN APPLICATION TO THE U.S. DEMAND FOR FACTORS OF PRODUCTION

Hendrik Wolff

University of Washington

February, 2009

Abstract

Behavioral assumptions are often the central building blocks of many economic theories, such as the assumption of profit or utility maximization. This paper proposes a procedure to test for behavioral assumptions. The test relies on the following principle: Behavior manifests itself in the form of "shape conditions." For example, if a firm minimizes costs, then, by standard microeconomic theory, the dual cost function is concave and increasing in input prices. The intuition of the proposed test is simple: if for a given dataset we statistically reject the shape properties, then we reject the underlying behavioural assumptions of the economic theory. To make the test work, we first estimate a flexible functional form without imposing the shape conditions. Secondly, we re-estimate the function subject to the shape conditions. Finally, the comparison of the restricted estimate to the unrestricted estimate provides the test statistic.

The challenging part of this test is the estimation of the restricted model. Literature has suggested several shape-imposing estimators, but it is not clear which of these to use in practice. In this paper, we apply a series of such estimators, and their comparison within this context provides several insights into the advantages and disadvantages.

Keywords: Demand System, Cost Function, Shape Restrictions, Nonlinear Inequality Constraints, Flexible Functional Forms, Metropolis-Hastings, Regularity Conditions, Hypothesis Test

Acknowledgement: I wish to thank Guido Imbens, Jeffrey LaFrance, Jenny Lanjouw, Arnold Zellner, Jen Brown, Ricardo Cavazos, Yanay Farja, and Muzhe Yang, for motivation, and valuable discussions during the course of this research. Furthermore, thanks are due Ernst Berndt, for providing the data set.

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

While a difficult literature, we believe that research on models permitting flexible imposition of true regularity should expand.

-- William A. Barnett & Meenakshi Pasupathy, 2003 --

1. Introduction

A lot of work in economics can be characterized as follows: researchers start by making a set of behavioral assumptions and developing a theoretical model. After deriving the functions of interest, we (econometrically) estimate the resulting system of equation(s), and finally, use the empirical model for policy analysis. The link between ‘economic theory’ and the ‘empirical model’ is the data. Two questions motivate this paper: is the data in line with the assumed behaviour implied by the economic theory, and how can we test for this relationship?

One critical implication of economic theory is that the estimating functions often have to satisfy ‘shape restrictions’. For example, if it is assumed that a firm exhibits behaviour of a cost minimizer (a standard assumption in many economic models derived from duality theory), then the dual cost function is concave and monotonically increasing in input prices and convex and increasing in output.¹ Rejecting these shape conditions is then equivalent to rejecting duality theory.

In applied work, the researchers’ aim is the consistency of the empirical model with the underlying economic theory. In fact, many functional forms represent small subsets of the class of functions that the economic theory generates. For example, the popular CES and Cobb-Douglas cost functions are both concave and increasing in prices (as required by standard micro-economic theory), but these functions are, at the same time, restrictive by fixing the elasticities of substitutions. Moreover, this does not allow formal testing of the underlying economic theory because these functions are themselves strictly within the class of functions generated by the theory. One way to step out of this box is to use flexible functional forms (FFF)—examples are the Translog, Generalized Leontief, and Symmetric McFadden and higher order expansions thereof.

¹ Also, by the Shephard Lemma, a dual cost function implies that the input demand system is downward sloping such that the law of demand holds and the marginal cost functions are increasing in both prices and output. Similar relationships between behavior and shape conditions hold in many other contexts: if individuals are utility maximizers, then the indirect utility function is quasi-convex in prices. If firms are profit maximizing, then the supply function is upward sloping and the profit function is convex in both output and input prices.

These functions can (locally) represent any relationship generated by the economic theory as well as exhibit shape properties not predicted by the economic theory.

The potential gap between a well-established economic theory, on one side, and the empirical model, on the other, is of great concern, as reflected in the large literature on regularity preserving estimation procedures produced in the past 30 years, see Gallant and Golub (1984), Diewert and Wales (1987) and Barnett and Binner (2004) for literature reviews on this topic. Nonetheless, in empirical applications theoretical assumptions are often violated and policy recommendations derived from such models are dubious at best; see the discussions in Salvanes and Tjøtta, 1998; Griffiths, O'Donnell and Tan-Cruz, 2000; Barnett, 2002; Blundell, 2004.²

2. Outline

This paper presents a framework for estimation and inference to test the fundamental behavioral assumptions, such as profit maximization or utility maximization, characterized by optimizing some objective function subject to constraints. The test relies on the following principle: Behavioral assumptions manifest themselves in the form of uniquely defined ‘shape conditions’. The intuition of the proposed test is simple: if we statistically reject the implied shape properties, we then reject the economic theory as well. Hence, the objective is to test the gap between the empirical model and the economic theory.

To make our test procedure work, we first estimate an FFF unrestrictedly. If it satisfies all the required shape restrictions, then the estimated empirical model does not reject the economic theory. If, however, the estimated model violates the shape restrictions, a second procedure follows: we re-estimate the function, subject to the side conditions, such that all shape conditions are satisfied. Finally, the comparison of the restricted estimate to the unrestricted estimate provides us with the test statistic.

In the literature on FFF, we seldom observe cases where the underlying theory is formally tested, and if at all, these tests are rather partial or ad hoc. For example, some papers report the percentage of data points at

² Monotonicity and convexity conditions are the most frequently violated. Curvature alone might be successfully implemented with quadratic functional forms but these may fail to produce correct monotonicity. For common parametric and semi-non-parametric procedures, these violations are the consequence of the fact that standard estimation methods can impose the shape restrictions partially only—as proven by Lau, 1986 for a general class of parametric models.

which they violate the shape conditions. This informal test statistic, however, is problematic. In fact, it turns out that a function violating the shape conditions at all data points can be very “close” to the economic theory. However, the other case is possible as well: a function violating a shape restriction at one singular data point can still imply an empirical model that is completely out of sync with economic theory. We argue that our procedure therefore outperforms such ad-hoc tests.

Clearly, the underlying principle (Likelihood ratio) of our proposed test is not new. Rather, we put together different pieces from the existing literature. The contribution of this paper lies in (i) pointing out that the previous methods within the FFF literature of testing for behavioral assumptions are not based on firm grounds, (ii) discuss and compare the performance of a set of estimators with the ability to impose the shape conditions, and (iii) illustrate the proposed procedure to an empirical data set.

For illustration we apply our method to the “Berndt and Wood” dataset—that has been intensively used to test the performance of new estimators in the econometric literature—estimating a flexible input demand system of four production factors to the U.S. Our empirical results demonstrate that the estimator choice can lead to significantly different policy implications. In particular, we find that estimates based on standard estimators could have erroneously rejected the duality hypothesis, whereas our preferred estimator provides strong support that the Berndt and Wood data is in fact consistent with economic duality theory.

The writing of this paper emphasizes the empirical issues relevant to the application. In reviewing the literature on shape imposing estimators, we find few arguments on *why* these methods require application. In the following section we discuss why this topic is important in applied economics. Sections 4 and 5 outline various regularity-imposing techniques, section 6 illustrates these approaches empirically, and in section 7 we provide conclusions.

3. Preliminaries

Although many will agree that a sound empirical model must be consistent with the underlying economic theory, doubt may yet exist as to the purpose of employing regularity-imposing estimators. If we do not allow data to speak for itself—but force it into relationships dictated by theory—does this not imply that the assumed economic theory is wrong, the model inaccurately specified, or that the data quality is not appropriate for the estimation?

Before offering an answer, we first review some of the concepts involved. In this paper we estimate a factor demand system which is derived by Shephard's Lemma from the dual cost function $c^*(p,y,t) = \min\{p \cdot x \mid f(x,t) \geq y\}$. Inputs are denoted by $x \in \mathfrak{R}_+^N$ and are transformed into output $y \in \mathfrak{R}_+$ by a smooth production function $f(x,t)$, which depends upon technological change $t \in \mathfrak{R}_+$. Duality theory implies that $c^*(z)$, where $z = [p,y,t] \in \mathfrak{R}_+^{2+N}$, is

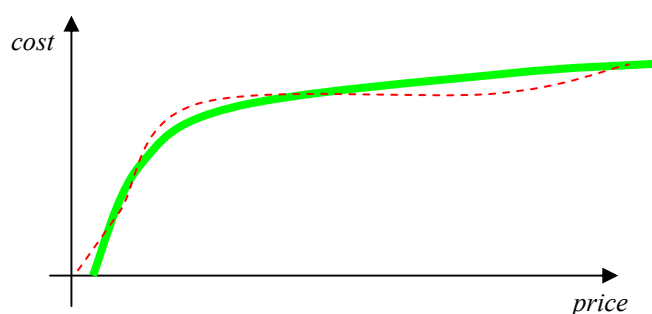
- HD1 : homogenous of degree one in input prices $p \in \mathfrak{R}_+^N$
- M_p : monotone increasing in p
- C_p : concave in p
- C_y : convex in y , and
- M_y : monotone increasing in y .

If a function satisfies properties HD1, $M_{p,y}$ and $C_{p,y}$, we then say that the function is 'regular' or synonymously, 'well-behaved.' To impose these regularity conditions, the literature for many years concentrated on the derivation and estimation of factor demand systems from globally regular generating functions, such as the Cobb-Douglas and the Constant Elasticity of Substitution.³ These *first order flexible functional forms* satisfy the restrictions of homogeneity, monotonicity, and curvature by well-known parametric restrictions; at the same time, these forms restrict the potential values for the second order effects prior to the estimation. This implies that the elasticities of substitutions (which are key parameters to reveal, as we have discussed in the introduction) cannot be estimated, but are fixed. Consequently, the literature in the

³ *Global, local and regional* describe properties of different subsets of the right hand side variable space \mathfrak{R}_+^{N+2} (here spanned by z). Whereas global refers to the unbounded domain \mathfrak{R}_+^{N+2} , local refers to one singular point or multiple singular points, and regional refers to a connected subset of \mathfrak{R}_+^{N+2} .

1970s moved toward local approximation functions to the true data generating function itself, by series expansions. The result was the class of *second order flexible functional forms*, such as the popular Translog, Generalized McFadden, and the Generalized Leontief, providing the capability to attain arbitrary elasticities of substitution. Nevertheless, this increased flexibility came with a cost: sacrifice of a guarantee of obtaining regularity.⁴ In fact, a series of studies (for a review, see Barnett, Geweke and Wolfe (1991)) demonstrated that these forms often have very small regions of theoretical regularity.

Fig. 1 : True versus approximation function



Even if the true data generating cost function c^* is deterministic and regular (bold - green), the estimated approximation function \hat{c} (dashed - red) can be irregular. The movement from c^* to \hat{c} is not a simple parallel movement. Instead, (because realized (weighted) residuals sum up to zero) \hat{c} oscillates around c^* . This phenomenon has been demonstrated with numerical examples by WHM using the Translog, the Generalized Leontief, and the AIM.

As previously noted by Moschini (1999), from a positive point of view, violations of the regularity conditions may call into question the applicability of the dual demand theory to a particular data set. This raises the issue of whether one should really force the function to satisfy economic theory, or if one should rather let the data speak and search for other (non-neoclassical) explanations. We argue that regularity-preserving techniques are indispensable for at least three reasons⁵:

- (a) Any finite order flexible functional form c represents an approximation to the true function c^* . If c^* is regular and *stochastic*, then \hat{c} , estimated with some non-regularity preserving estimator, can fit outliers produced by c^* and thus violate regularity.
- (b) The upshot is, even if c^* was regular and *deterministic*, \hat{c} can oscillate around the true relationship. Because of its approximating nature, \hat{c} has a different tracking behavior over its domain, so it does not lie completely

⁴ For linear-in-the-parameters functional forms, Lau (1986) proved flexibility incompatible with global regularity with the imposition of both concavity and monotonicity. 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 first order series expansion, the Cobb-Douglas, which has constant cross elasticities of value one.

⁵ Empirical evidence demonstrated that economic theory matters: A ‘regular model’ may often forecast better out of sample - although their ‘in sample’ fit statistics are inferior compared to an irregular model. This is an interesting point but not a general result. For a recent discussion on this, see Edwards and Terrell (2004).

above c^* , but slightly next to it, as shown in fig. 1. This is perhaps the most important reason for the use of regularity-retaining techniques in practice. Otherwise one risks erroneously concluding that some data is ill behaved, whereas, in fact, the true data generation process is regular.

(c) A regularity preserving point estimate is required for correctly specifying hypothesis tests. More on this issue outlined in the empirical section.

Having motivated the need for shape imposing techniques, the next section reviews a series of estimators that are currently available to economists. The performance of these estimators undergoes evaluation in section 6.

4. Shape imposing estimators

The focus of this section is on methods to impose the inequality constraints to obtain $M_{p,y}$ and $C_{p,y}$. Such methods can be categorized into three groups: (a) global, (b) local and (c) regional imposition of regularity (for definitions of these terms see footnote 2). Global and local approaches are by far the most common methods currently employed by economists. Our proposed method, following ideas by Gallant and Golub (1984) and Terrell (1996), instead pursues the regional approach. Through this process, conditions are imposed on a connected subset ψ of the domain of the function being estimated. The connected subset represents what we call the *empirically relevant region*, and is defined by the model analyst (see Definition 1 below). In our view this *regional approach* offers important advantages over the *local* approach because it imposes theoretical consistency not only locally, at a given singular 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, generally leading to a better model fit to the sample data compared to the *global* imposition of regularity. In the empirical section 6 we test for these claims, comparing local, regional and global approaches.

This regional regularity approach first proposed by Gallant and Golub (1984) has an advantage in that flexibility of the functional form can be maintained to a large degree while staying theoretically consistent in the region where inferences will be drawn. In addition, imposing regional regularity generally leads to a better statistical fit of the data to the model, compared to the global regularity approach. However, Gallant and Golub

did not demonstrate the tractability of this approach and it seems that empirical implementation can be formidable with optimization tools currently available.

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

In the next section we propose an estimator that substantially mitigates previous difficulties and inconsistencies in applying the regional regularity concept. Here, while referring readers to Wolff, Heckelei and Mittelhammer, 2006 (in the following referred to as WHM) for the theoretical matters, we only briefly describe the main steps involved.

5. Regional regularity: An alternative

5.1. Definition of ψ and constraints to be imposed on ψ

The proposed technique is probably best explained through our empirical application of section 6 with $N = 4$ inputs and $T = 25$ observations. The two curvature conditions $C_{y,p}$ and the two monotonicity conditions $M_{y,p}$

must hold on a connected subset $\Psi \subset \mathfrak{R}_+^{2+N}$ of the price \times output \times time space. $C_{y,p}$ and $M_{y,p}$ can be characterized by $H=4$ vector-valued-functions $i_h(\mathbf{z};\boldsymbol{\beta})$, $h=1,\dots,H$, whereby the restrictions hold whenever, for a given $\boldsymbol{\beta}$, \mathbf{i} is nonnegative for all z in the relevant region Ψ ,

$$\mathbf{i}(\mathbf{z};\boldsymbol{\beta}) \equiv [i_1, i_2, \dots, i_H] \geq \mathbf{0} \quad \forall p, y, t \in \Psi.$$

Hence for monotonicity and curvature we define the following four sets of constraints:

$$i_1 = \nabla_p c \geq 0, \quad i_2 = \nabla_y c \geq 0, \quad i_3 = -\text{eig}[\nabla_{pp}c] \geq 0, \quad \text{and} \quad i_4 = \text{eig}[\nabla_{yy}c] \geq 0.$$

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 conditions *regionally*. For this purpose we now define a particularly relevant Ψ .

Definition 1: *The set Ψ is a closed and connected subset of \mathfrak{R}_+^{2+N} that covers the empirically relevant region, defined as containing all T sample observation as well as any S regressor points that will be used for subsequent analyses and/or simulations based on the estimated model.*

Note that conceptually previous global and local approaches only differ in the way Ψ is defined. If $\mathbf{i} \geq \mathbf{0} \quad \forall z \in \Psi$, we say that regularity is imposed (i) *locally* if Ψ consists of one or more singular disconnected points, (ii) *globally* if $\Psi = \mathfrak{R}_+^{2+N}$, and (iii) *regionally* if Ψ is some connected subset of \mathfrak{R}_+^{2+N} .

Before proceeding, let us restate the importance of definition 1. 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. We argue that once it is ensured that the empirically relevant set Ψ 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. Whereas the functions shape properties like $M_{p,y}$ and $C_{p,y}$ on Ψ are purely determined by economic theory, the coordinates of Ψ (within the right hand side variable space \mathfrak{R}_+^{2+N}) have to be defined by the analyst having in mind the purpose of the model, hence knowing about the $T + S$ data points, i.e. at which ranges one aims to make inferences.

5.2. Bayesian framework and numerical integration

In Bayesian econometrics the distribution of interest is the regularity posterior $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$, which in our case depends on $\boldsymbol{\psi}$.⁶ We now turn towards the simulation technique used to generate outcomes from $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$, which are then used to obtain point estimates and to draw posterior inferences. Based on Griffiths, O'Donnell and Tan-Cruz (2000) the Metropolis-Hastings Accept Reject Algorithm (MHARA) is used to generate J (pseudo-) random outcomes, $\mathbf{b}^{(j)}, j = 1, \dots, J$ from $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi})$ on the regular support $\Theta^R|\boldsymbol{\psi} = \{\boldsymbol{\beta}: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \forall \mathbf{z} \in \boldsymbol{\psi}\}$. To account for the regularity prior $\mathbb{1}\{\boldsymbol{\beta} \in \Theta^R|\boldsymbol{\psi}\}$, the simulator should ensure that any drawn parameter vector $\mathbf{b}^{(j)}$ implies regularity of $\mathbf{c}(\mathbf{z};\mathbf{b}^{(j)})$ for every point \mathbf{z} in the predefined set $\boldsymbol{\psi}$, i.e. $\mathbf{b}^{(j)} \in \Theta^R|\boldsymbol{\psi} \forall j$. Since there are an infinite number of points in $\boldsymbol{\psi}$, they cannot all be checked explicitly. The connectedness is approximated by a fine grid, denoted by the disconnected set $\boldsymbol{\psi}_g \subset \boldsymbol{\psi}$ which possibly consists of tens-of-millions of equidistant distinct points. Within the Metropolis Hastings chain an additional Accept-Reject algorithm is implemented to guarantee that $\forall \mathbf{b}^{(j)}$ the regularity conditions hold for any single grid point. This implies that $\mathbf{b}^{(j)} \in \Theta^R|\boldsymbol{\psi}_g \forall j$, whereby $\Theta^R|\boldsymbol{\psi}_g$ is the *approximated* regularity posterior support, which will tend toward the actual set $\Theta^R|\boldsymbol{\psi}$ the finer the approximation grid $\boldsymbol{\psi}_g$.

5.3. Approximating $\boldsymbol{\psi}$

Key to the success of empirical implementation of the technique is to have ‘good’ construction procedures for $\boldsymbol{\psi}_g$. The design of $\boldsymbol{\psi}_g$ not only influences our confidence in the approximation but also determines the tractability of the method since a high number of grid points increases computing time considerably; too few grid points, however, will raise concerns about potential violations of the regularity conditions in $\boldsymbol{\psi} \setminus \boldsymbol{\psi}_g$. To begin the discussion we define (as in Terrell, 1996) $\boldsymbol{\psi}^\square$ as a hypercube (the superscript \square refers to the cube approach): Let $z_i(\boldsymbol{\psi}_{\min})$ and $z_i(\boldsymbol{\psi}_{\max})$ represent the minimum and maximum of the i -th right hand side variable. The grid is constructed by selecting $F = 10$ equidistant values for each variable: $z_i^f = z_i(\boldsymbol{\psi}_{\min})$

⁶ Let Θ be the K -dimensional parameter space. If all regularity conditions hold for all values of \mathbf{z} in $\boldsymbol{\psi}$, the regular parameter set is defined as $\Theta^R|\boldsymbol{\psi} = \{\boldsymbol{\beta} \in \Theta: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0} \forall \mathbf{z} \in \boldsymbol{\psi}\}$, hence $\Theta^R|\boldsymbol{\psi}$ is dependent on the choice of $\boldsymbol{\psi}$. The marginal prior on $\boldsymbol{\beta}$ is specified as an indicator function $p(\boldsymbol{\beta}|\boldsymbol{\psi}) = \mathbb{1}\{\boldsymbol{\beta} \in \Theta^R|\boldsymbol{\psi}\}$ where the prior equals 1 if regularity holds at the value $\boldsymbol{\beta} \forall \mathbf{z} \in \boldsymbol{\psi}$, and equals 0 otherwise. Throughout the paper we assume the standard ignorance prior for the $N \times N$ covariance matrix $|\boldsymbol{\Sigma}|^{-(N+1)/2}$. The posterior distribution for $\boldsymbol{\beta}$ is then derived by applying Bayes rule, $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto L(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y}) \cdot \mathbb{1}\{\boldsymbol{\beta} \in \Theta^R|\boldsymbol{\psi}\} \cdot |\boldsymbol{\Sigma}|^{-(N+1)/2} d\boldsymbol{\Sigma}$, where $L(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})$ is the normal likelihood function.

+ $(f-1)F^{-1}(z_f(\boldsymbol{\Psi}_{\max})-z_f(\boldsymbol{\Psi}_{\min})) \forall f \in \{1,2,\dots, F\}$ and using all possible $Q = 10^6 = F^{\dim(\boldsymbol{z})}$ combinations to generate the Q -grid $\boldsymbol{\Psi}_g \subset \boldsymbol{\Psi}^\square$. In order to circumvent the approximate nature of this representation, WHM identified conditions under which checking a certain key point in $\boldsymbol{\Psi}^\square$ guarantees regularity in well defined neighborhood. The purpose is then to find a collection of such key points that guarantees overall regularity $\forall \boldsymbol{z} \in \boldsymbol{\Psi}^\square$. These procedures are described in detail in WHM and implemented in the application below, leading to a reduction in regularity checks to a total of $Q^* = 343900 < Q = 10^6$. Notably, the new Q^* -grid, while improving the computational speed of the algorithm, maintains the same accuracy of approximation obtained from the original Q -grid.

This reduction of regularity checks can be improved; up until this point the literature on regional regularity defined $\boldsymbol{\Psi}$ as one convex set $\boldsymbol{\Psi}^\square$. (Convexity had been originally a requirement for the constraint optimization program by Gallant and Golub, 1984.) We relax this assumption and let $\boldsymbol{\Psi}$ to be any connected set satisfying Definition 1. By constructing a nonconvex set, while maintaining the same accuracy of the approximation obtained from the original 10^6 grid points, the number of regularity checks is reduced to occur, for example, at $316 = 1+(F-1)(T+S)$ grid points only, if the number of out of sample forecasts $S=10$. These grid construction rules lead to enormous cutbacks in computing time, and they enhance the tractability of regional regularity preserving estimators.

5.4 Point estimates and the relation to Maximum Simulated Normal Likelihood

As far as we are aware, all previous studies define the point estimate as the mean $E[\boldsymbol{\beta}]$ of the regularity posterior $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\Psi})$. This may result in regularity violations because in general $\Theta^R|\boldsymbol{\Psi}$ is not a convex set. Instead using the mode $\boldsymbol{\beta}^{(\text{mode})} = \arg \max_{\boldsymbol{\beta} \in \Theta^R|\boldsymbol{\Psi}_g} \{p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\Psi}_g)\}$ guarantees that the point estimate resides in the regular support $\Theta^R|\boldsymbol{\Psi}$. In order to approximate the solution based upon the MCMC outcomes $\{\mathbf{b}^{(j)}\}_{j=1}^J$, one can simply compare the values $p_u(\mathbf{b}^{(j)}|\mathbf{y}) \forall j$ resulting from the MHARA as $\mathbf{b}^{(\text{mode})} = \arg \max_{\mathbf{b}^{(j)}} \{ |(N-L)\boldsymbol{\Sigma}(\mathbf{b}^{(j)})|^{-N/2} \}$. This point estimate is used in the application below.

The proposed technique can be applied to the Bayesian and to the Classical frameworks. In the

Classical framework one would maximize a likelihood function subject to the inequality constraints and the numerical point estimate of the maximum simulated likelihood is the mode. This Classical mode is exactly identical to the above defined Bayesian point estimate $\beta^{(\text{mode})}$ if, as we have done above, an uninformative prior distribution on $\Theta^R|\psi$ is employed (see footnote 5). Here we prefer the Bayesian interpretation because finite sample confidence intervals and standard errors of functions of β can be directly computed with the MCMC draws. Instead deriving the Classical distributions could be tremendously challenging and in general requires more time intensive numerical procedures (like bootstrapping). The computational burden is mainly due to the many inequality constraints $\mathbf{i} \geq \mathbf{0}$ which have to hold for all $z \in \psi$. Also since $\beta^{(\text{mode})}$ could lie on the boundary of $\Theta^R|\psi$ further complications arise (see Geweke 1986, Andrews 1999, 2001).

6. Empirical Illustration

Background to the Application of Estimating Input Demand Systems

An motivating application exemplifies the use of the demand system: Climate change concerns drive many countries to debate over imposing a tax on energy use intending to reduce CO₂ emissions. To quantitatively assess the costs and benefits of such a policy, an analyst requires two information: the own price elasticity of demand and secondly, the cross-price elasticities that describe the effects on important markets that are linked to energy; in fact with the tax policy in place, firms could substitute away from energy towards other inputs such as capital and labor—which may be less polluting but more costly. To this end, estimated demand system have provided key ingredients to answer many important questions in production analysis (Chambers 1988, Griffiths, O’Donell and Tan Cruz 2000, Kumbhakar and Tsionas 2005), policy studies and welfare analysis (Evans and Heckman 1984, 1986, Koebel, Falk and Laisney 2003), as well as in the debate on the sources of economic growth (Mankiw, Romer and Weil 1992, Hsieh 2000, Antras 2004).

Our empirical illustration contains 3 subsections: In section 6.1, we re-estimate the demand system for four production inputs to the U.S. manufacturing sector using the Berndt and Wood (1975) data set for capital (K), labor (L), energy (E), and materials (M). For this data set it has been reported to violate the regularity

conditions implied by economic theory. For that reason the KLEM data has been applied to a considerable number of regularity imposing techniques providing a substantial basis on which to investigate the performance of these alternative estimators (see among others Berndt and Wood 1975, Berndt and Khaled 1979, Galant and Golub 1984, Diewert and Wales 1987, Barnett, Geweke and Wolfe 1991, Friesen 1992, Terrell 1996).⁷

But one can ask if these theoretical notions have any practical importance. For instance, does the MCMC methodology produce elasticity estimates that are significantly different from the above other estimation approaches? Section 6.2 looks at the elasticity of substitution between capital and energy, a parameter that has attracted a great deal of attention in the last decades (see e.g. Apostolakis 1990). Finally, the last section goes one step beyond and asks of whether duality theory is appropriate for modelling the U.S. input demand system. Simply forcing the empirical model to be consistent with economic theory (by employing shape imposing estimators) could produce misleading results. Instead formal hypothesis tests should be carried out.⁸

The Berndt and Wood data have been described in more detail in many places in the literature (e.g. Berndt and Wood (1975), Berndt and Khaled (1979), Gallant and Golub (1984)). Table 1 provides summary statistics of the annual data from 1947 to 1971. In particular the min/max values of the right hand side price variables will be considered below for the construction of the various ψ sets.

⁷ In fact, it may even be one of the most frequently estimated input demand systems in the entire econometrics literature.

⁸ This clearly is of concern when investigating such important issues as the evaluation of the costs and benefits of an environmental tax. As a consequence of a 10% increase in energy price, calculated differences of the sectors ability for capital accumulation range in the magnitude of hundreds of million U.S. dollars, depending on which estimation approach is chosen. These numbers show that the impact on policy forecasts could be disastrous if regularity conditions are violated or the assumed substitution pattern is mistaken. Of further practical interest are the results that energy is a substitute to all inputs except capital and that energy demand is inelastic with the own price elasticity significantly about -.67.

Table 1: Summary statistics of the Berndt and Wood data set from 1947 to 1971

	input quantities				input prices				output
	K	L	E	M	ρ_K	ρ_L	ρ_E	ρ_M	y
mean	20.45	106.07	16.78	237.56	1.18	1.77	1.35	1.30	313.80
std	7.77	43.59	5.54	85.14	0.19	0.46	0.12	0.14	87.67
min	8.58	45.10	7.76	112.35	0.74	1.00	1.00	1.00	182.83
max	34.11	190.26	29.48	407.71	1.50	2.76	1.65	1.55	466.83

Variables are produced using index numbers and deflators. For details on the data construction see Berndt and Wood (1975) and Berndt and Khaled (1979).

6.1. Comparing Shape Imposing Techniques – An Illustration using the Berndt and Wood Data

The main purpose of the following eight sets of estimations is to assess potential advantages of the regional approaches compared to the local and the global approaches both in terms of model fit and the propensity for regularity violations. Performance statistics of various estimators as applied to the second order flexible Generalized Leontief cost function

$$c(\mathbf{z}; \boldsymbol{\beta}) = \sum_{i=1}^N \sum_{j=1}^N b_{ij} p_i^{0.5} p_j^{0.5} + \sum_{i=1}^N b_i p_i + \sum_{i=1}^N b_{it} p_i t y + t \sum_{i=1}^N a_i p_i + y^2 \sum_{i=1}^N \beta_i p_i + y t^2 \sum_{i=1}^N \gamma_i p_i$$

with $b_{ij} = b_{ji}$, are displayed in table 2. In order to be able to directly compare our results with previous studies, we use the exact same specification of the demand system as in Diewert and Wales, 1987 and in Terrell, 1996.⁹ Hence the N estimated equations are

$$\mathbf{x}/\mathbf{y} = \nabla_p c(\mathbf{z}; \boldsymbol{\beta})/\mathbf{y} + \mathbf{u}. \quad (1)$$

It is assumed that the $T \times 1$ error vectors \mathbf{u}_n , $n = 1, \dots, 4$ are contemporaneously correlated, such that the estimating equations can be written in form of the Gaussian seemingly unrelated regression (SUR) system.

Finally $t = 1, 2, \dots, T$. For details on the specification see Diewert and Wales, 1987 and Terrell, 1996.

Duality theory restricts the Generalized Leontief cost function to be M_p , M_y , C_p and C_y . In general, economists are well aware of these fundamental relations, imposing *all* four of these conditions when estimating first order flexible functional forms. In contrast, the standard practice is that only a small subset of

⁹ For a motivation of this particular specification see Diewert and Wales (1987) and Terrell (1996). In particular, all right hand side variables are assumed to be exogenous. This seems to be a standard approach in this literature: Except for the Berndt et al. articles in the seventies, none of the above papers estimating the KLEM input demand system uses instruments. For a justification of this, see the recent discussions by Diewert (2004), Barnett and Binner (2004), and Antras (2004). Another possible extension is to estimate the system in the context of an error correction model (Friesen, 1992). For both 3SLS and the ECM specification it is straightforward to impose shape conditions by using Terrell's Gibbs and the MHARA simulator.

these conditions is enforced when using the Generalized Leontief or any other second order flexible forms. In particular, the three conditions $M_{p,y}$ and C_y conditions have rarely been considered. A remarkable exception of a paper explicitly imposing both C_p and M_p is Terrell 1996. His results will be shown below. But still, violations of the other restrictions M_y and C_y lead to serious flaws, such that, for example, with rising output, ceteris paribus, the total cost of production decreases.

The fact that the literature ((Berndt and Wood 1975, Berndt and Khaled 1979, Galant and Golub 1984, Diewert and Wales 1987, Berndt, Geweke and Wolfe 1991, Friesen 1992, Terrell 1996) contemplated subsets of the regularity conditions cannot be justified from a perspective of economic theory. Why should a violation of monotonicity be less harmful than a concavity violation? This development might only be explained by the lack of estimators that have the ability to maintain *overall regularity*. Such a gap between economic theory and the empirical model is problematic for the interpretability of the results and especially worrisome if one wishes to derive any policy conclusions for the U.S. manufacturing sector, which accounts for about 20% of the GDP. To our knowledge for this input demand system this paper represents the first study systematically taking into account *overall* regularity when using flexible functional forms.

Table 2: Regularity imposing sets

Approach		Definition of ψ	Comment
Unconstrained		$\psi = \emptyset$:	$M_{p,y}$ and $C_{p,y}$ is not imposed. ‘Unconstrained’ refers to the ‘inequality constraints’ only: we do impose symmetry and HD1 by parametric equality restrictions.
Local		$\psi = z_1$	We use the first observation in the sample (1947) because this represents our results comparably to the tables in Berndt and Wood (1979), Diewert and Wales (1984), Barnett et al., (1991) and Terrell (1996). Imposing the regularity conditions at other points in the sample space does not essentially change the empirical results.
Global		$\psi = \mathfrak{R}_+^{2+N}$	Nonnegative orthant of all right hand side variables of z .
Regional	Cube ¹⁰ approach	$\Psi^{cp}_1 = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^3 [1.0, 1.5]\}$	This was chosen by Terrell (1996). It does not cover the entire empirical relevant data space. Some observed prices lie outside the [1.0, 1.5] interval, see min/max values in table 1.*
		$\Psi^{cp}_2 = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^3 [0.5, 3.0]\}$	This set was chosen by Terrell (1996). It ensures that Ψ^{cp}_2 covers all observed prices and beyond.*
		$\Psi^{cz}_i = \{z \in \Psi^{cp}_i \times [y_{\min}, y_{\max}] \times [t_{\min}, t_{\max}]\}$	$\Psi^{cz}_i, i \in \{1,2\}$: These sets expand Ψ^{cp}_i over the remaining dimensions t and y .
	String approach	$\psi^{\text{string}} = \bigcup_{i=1}^{26} \psi_i$	ψ^{string} covers $26 = T+1$ points in \mathfrak{R}_+^{2+N} by connecting straight lines ψ_i between the right hand side variables mean, \mathbf{z}^M , and each of the T observations. We approximate each line ψ_i by ψ_{ig} by taking 10 equidistant grid points between \mathbf{p}^M and the i^{th} observation z_i , leading to $1+(F-1)T = 226$ grid points.

* Set is Lebesgue measure zero because it does not expand into the dimension y and t .

Columnwise table 3 displays the various approaches imposing restrictions on differently designed sets ψ , which definitions are first provided in table 2. In Table 3, the symbols in parentheses indicate which shape conditions are imposed. For example, $\Psi^{cp}_2(M_p, C_p)$ implies that the estimator imposes concavity and monotonicity with respect to p . If all four regularity conditions are imposed, we simplify notation to (‘all’). For each approach we display the results as applied to both point estimates, the mean of the regularity posterior distribution, and the mode. Ex post, as in Terrell, after each estimation, the regularity conditions are evaluated at empirical relevant sets Ψ^{\square}_1 and Ψ^{\square}_2 as indicated in the respective rows. Violations are expressed as the percentage of grid points where violations occur, whereby the grids are constructed as outlined in above section 5.3.

¹⁰ In this paper all grid sets ψ^{\square}_{ig} are constructed with $F = 10$, as in Terrell, 1996. As pointed out by a referee a constant F is not unproblematic since the likelihood of regularity violations at non-grid points varies as the cube’s volume V changes. Here V increases dramatically from $V(\Psi^{cp}_1) = 0.0625$ to $V(\Psi^{cp}_2) = 150.0625$. As an alternative one could express the number of grid points as a function of the volume, $F(V)$.

Table 3: Generalized Leontief Input Demand System, estimated with 8 different approaches

Model Performance Statistics: Regularity Violations / Fit Statistics		Estimation Approach															
		Unrestricted	Global			Local		Regional									
		1	2a	2b	3	4	5	6	7	8							
Restricted domain of the function (and in paranthesis which shape conditions are imposed)		\emptyset (none)	$\mathcal{S}^{2+N}(C_p)$	\mathcal{S}^{2+N} (all)	z_1 (all)	$\psi^{\text{op}}_1(C_p, M_p)$	$\psi^{\text{op}}_2(C_p, M_p)$	ψ^{cs}_1 (all)	ψ^{cs}_2 (all)	ψ^{string} (all)							
evaluated at	% of shape violations at grid points	mode	mean	mode	mode (\mathbf{h}^{est})	mean	mode	mean	mode	mean	mode	mean	mode	mean	mode	mean	mode
ψ^{op}_1	C_p violations	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	M_p violations	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C_p violations	4.0	25.1	33.4	0.0	0.0	0.0	0.0	0.0	0.0	53.0	0.0	0.0	0.0	0.0	0.0	0.0
	M_p violations	0.0	2.0	11.3	0.0	0.0	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ψ^{op}_2	C_p violations	100.0	0.0	0.0	0.0	27.1	27.1	16.1	27.0	0.0	0.0	15.2	27.0	2.2	0.0	27.4	27.4
	M_p violations	3.1	45.8	0.0	0.0	5.0	5.0	5.2	5.3	0.0	0.0	3.6	5.3	0.0	0.0	4.9	5.8
	C_p violations	31.2	33.4	25.1	0.0	32.9	32.9	35.4	33.3	29.7	50.5	1.6	33.3	0.0	0.0	0.0	32.8
	M_p violations	1.3	11.3	2.0	0.0	1.8	1.8	1.7	1.9	1.8	0.9	2.4	1.9	0.0	0.0	4.8	2.4
Generalized Variance of the Fit		1.44	0.52	0.52	0.27	1.26	1.26	1.18	1.26	0.74	0.76	1.14	1.26	0.17	0.27	1.20	1.26

The model fit is calculated by the *Generalized Variance of the Fit*. See e.g. Barnett (1976). Here it is defined as $100 \cdot |\Sigma|^{-1}$.

The statistic is proportional to the ordinate value of the unconstrained posterior $p(\mathbf{h}|y, \mathbf{O})$ evaluated at the respective point estimate.

Due to the choice of priors, here this measure is proportional to the likelihood value of the unconstrained Maximum Normal Likelihood regression.

6.1.1 Unconstrained estimation

In the first column of table 3 we estimate the demand system by iterated SUR unrestrictedly. Firstly, compared to any other columns, the unrestricted estimate \mathbf{b}^u provides the best model fit statistics but $c(z; \mathbf{b}^u)$ violates the regularity conditions *everywhere* in Ψ^{\square} leading to, among other things, a failure of the law of demand.¹¹ Contemplating these poor regression results, a researcher could pursue a multitude of directions, until something more consistent is obtained, i.e. trying other functional forms or applying the data to another economic theory which might be less demanding in terms of the regularity conditions. If goalposts are changed however in an ad hoc manner, such procedures can be rife with statistical testing and verification problems. Bearing in mind the discussion in section 3, the fact that c heavily violates the regularity conditions does not necessarily imply that Generalized Leontief functional form is inappropriate or that the duality theory has to be rejected. Hence other estimation approaches leading to a well-behaved economic model are required to test this hypothesis. (More on the hypothesis testing, see section 6.3).

6.1.2. Global Approach

The global approach to impose concavity is probably one of the most common techniques, when estimating flexible input demand systems. In the case of the Generalized Leontief, this unfortunately allows the cost function to model substitutes only (Diewert and Wales 1987). Because in the KLEM data at least energy and capital seem to have stark complementary relations, maintaining global concavity reduces the model fit. Moreover, the global imposition of concavity is not sufficient for *overall* regularity. In fact, all remaining conditions are violated as can be seen in column 2a. Now, applying the global approach not only to C_p but to all regularity conditions (estimated by (1) with $b_{ij}^u=0, i \neq j$) a-priori fixes the elasticity of substitution estimates to 0 (see footnote 3). Here, as displayed in column 2b, such as procedure performs very poorly in terms of model fit because this globally regular estimate \mathbf{b}^{gr} emerges from the ‘small’ parameter subset $\Theta^{\text{R}}|\mathfrak{R}_+^{2+N} \subset \Theta^{\text{R}}|\Psi$. In conclusion our empirical results indicate that the global approaches are extremely restrictive, a finding that agrees with the results in Diewert and Wales, 1987, Terrell, 1996 and the simulation study in WHM.

¹¹ For comparison, this estimation exactly repeats the unrestricted estimation of Diewert and Wales (1987: table II) and Terrell’s (table 3: 1996).

6.1.3. Local approach

The third column displays results from local imposition of overall regularity. As expected, here model fit is inferior when compared to the unrestricted approach, but is superior to the global approach. Unfortunately, the local approach does not guarantee that regularity is satisfied in all of the relevant empirical areas leading to very high percentages of violations at Ψ^{\square}_2 . One has to be careful, however, with the interpretation of these statistics on “percentage of violations at grid points”: On one side, violating shape conditions at 100% of the data points does not necessarily imply that the estimated model is “very far” from a good model. And on the reverse, a single violation of one grid point could lead to extremely poor results. This latter point is illustrated in Fig. 2 (this will be further discussed in section 6.3). Also note that Ψ^{\square}_1 does not cover all the relevant data points (compare the min/max values in table 1 and the definition of Ψ^{\square}_1 in table 2).

Since neither the unconstrained, local nor the global approach can produce a well-behaved flexible demand system of interest, we now turn to the regional approach.

Fig. 2: Illustration of an irregular cost function violating the shape restriction at one grid point

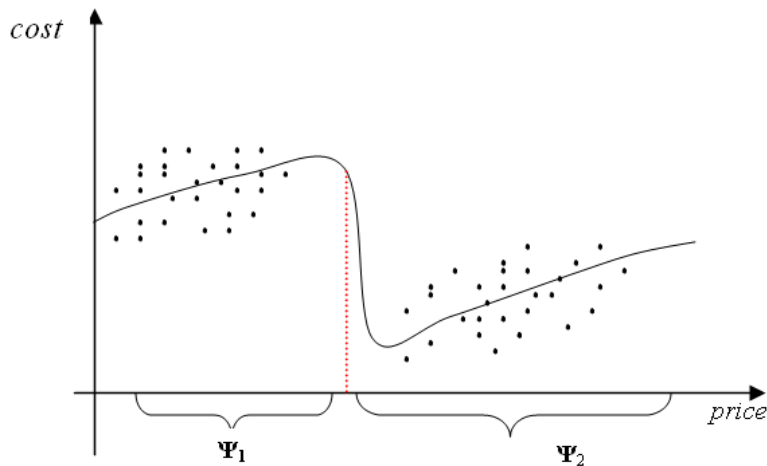


Illustration of a cost function partially regular within Ψ_1 and within Ψ_2 . The overall model is irregular on the domain $\Psi = \Psi_1 \cup \Psi_2$ although only one grid point is violated (here indicated on the domain by the red ray). Overall regularity is violated because cost must not decrease with rising input prices. Even in less dramatic circumstances – such as having softly oscillating functions – evaluating demand statistics at any violated grid point can lead to nonsensical results. At the non-concave areas in Fig. 1, for example, the estimated own price elasticity has the wrong sign.

6.1.4. Regional Regularity

We first replicate Terrell’s (1996) estimation, i.e. using the Gibbs accept reject simulator and the same Ψ -sets. Terrell applied this method to impose M_p and C_p . This successfully leads to regularity preserving results if interested in the function’s domain Ψ^{\square}_1 . In contrast, on the domain Ψ^{\square}_2 , the constraints with respect to y are violated (see column 4 and 5 in table 3).

We now turn towards the estimation method described in section 5 using the MHARA simulator. This method effectively imposes all regularity conditions on any set Ψ of interest. Hence, as shown in the columns 6 and 7 regularity holds in Ψ^{\square}_1 and Ψ^{\square}_2 . We display the results for both, the mean and the mode of the posterior. Interestingly, in column 7 the mean estimate \mathbf{b}^{mean} violates C_p , although C_p had been imposed! The reason that $\mathbf{b}^{\text{mean}} \notin \Theta^{\text{R}}|\Psi^{\square}_2$ is due to the fact that the regular parameter support $\Theta^{\text{R}}|\Psi^{\square}_2$ is not a convex set and demonstrates the importance of the proposition outlined in section 5.4. Instead our preferred point estimate, the mode $\mathbf{b}^{\text{mode}} \in \Theta^{\text{R}}|\Psi^{\square}_2$ and it guarantees regional regularity within the domain Ψ^{\square}_2 . In case of column 6, the mode increases the model fit (as measured by the likelihood value) by over 10% (from 1.14×10^{-2} to 1.26×10^{-2}). The percentage increase is even more dramatic in the case of imposing the regularity conditions on the larger set Ψ^{\square}_2 , in column 7, achieving an increase in model fit of 43.6%.

So far, we only have described the approaches based on a convex cube Ψ^{\square} . A motivation for the KLEM data set to investigate in non convex sets for Ψ is probably best described by Gallant and Golub, 1984: *‘The exogenous variable $[p_t$ and $y_t]$ for $t = 1947, \dots, 1971$ lie in a five dimensional space and can be projected into a three dimensional space with a negligible loss of information...The projected point cloud has an irregular shape. It is a sort of a fat rope lying mostly on the ground in the shape of a tilde (\sim) with the two endpoints and the middle elevated.’* This description indicates that constructing Ψ^{\square} as a convex cube including all the data points may lead to an unnecessary voluminous set containing relatively little data information. A simple construction rule of a nonconvex set containing all the KLEM data points is described in Table 2 and labelled as the ‘string approach’.¹² Comparing the posterior mean of the string approach versus the mean of

¹² It is straightforward to incorporate other out of sample areas into the construction of Ψ , i.e. in order to obtain more general

column 6 implies that here MHARA samples from a regular parameter superset $\Theta^R|\psi^{\text{string}} \supset \Theta^R|\psi^{\square}$, potentially benefiting the flexibility of the functional form. Finally, note that comparing the mean of the string approach to the mode leads to an improvement of the model fit of about 5.3%.

We particularly like to work with the string approach because of its attractive features. (a) It represents the method which is to the largest possible extent ‘data driven’, (b) it leads to a well-behaved demand model and (c) with 226 regularity checks, it is computationally *much* faster than the regional regularity preserving cube method, that checks one million times, as we have discussed in section 5.3.¹³

In table 3 we also supply performance statistics for the various approaches evaluated over a set which is larger than the set on which regularity was imposed. We do not necessarily advocate such an approach (i.e. defining ψ on a subset of the region where subsequent inferences will be drawn). Rather we include these results 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. This 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. We argue that once it is ensured that the empirically relevant 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.

Summarizing table 3, unconstrained and local regularity estimates increase the model fit in all specifications at the cost of violating regularity within ψ . This produces estimation results that are problematic in terms of economic interpretation and further analysis. Imposing regional regularity solves this problem and significantly increases the model fit when compared to the global approach. Moreover, apart from its appealing regularity preserving property, it is relevant for model fit to use the mode instead of the mean. Finally, the new proposed technique reduced computing time significantly, making the regional regularity approach more

domains relevant for forecasting purposes. See WHM for details.

¹³ With the Q^* -grid construction approach we need about 30% of the original Q -grid computing time (when imposing the shape conditions on the more dense Q -grid). In comparison the new string approach estimation is much faster requiring less than 0.05% of the original computing time.

tractable for empirical analysis. Instead of the full evaluation grid consisting of over one million points, only a maximum of 343900 points had to be evaluated only for the cube approaches. Furthermore, for the string approach only 226 points had to be assessed. This significantly decreased the computational burden when compared to previous approaches.

We have identified effective tools to generate well behaved input demand models. Initially, this may appear to be of interest to econometricians or economic theorists only. The imposition of regularity conditions, however, also leads to noticeable changes in own and cross price elasticities estimates, parameters that are of immense interest to a wide range of applied economists. These changes are investigated in the following subsection.

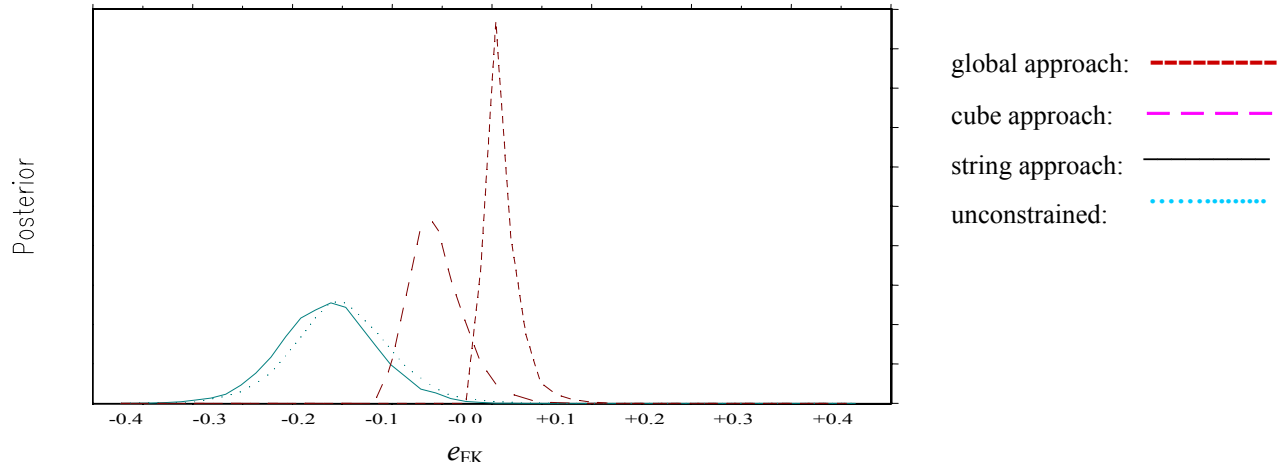
6.2. Elasticities

Further insights into the effects of imposing shape restrictions can be gained by examining estimated marginal posterior distributions for input demand elasticities $\partial x_i / \partial p_j \cdot x_i / p_j$. Table 3 reports the means, modes and standard deviations of these by MHARA simulated density functions. For the purpose of analyzing the potential effects of an environmental tax on energy use, here we are interested in the capital energy elasticity. The long-run growth potential of the manufacturing sector depends crucially on the magnitude of this parameter (see Dasgupta and Heal (1979), Chapter 6). In particular the question of whether capital and energy are complements or substitutes has received a great deal of attention (see e.g. Apostolakis 1990). If they are substitutes, then an increase in energy taxes would lead, *ceteris paribus*, to an increase in the capital stock, potentially benefiting the sector in the long run. In this case, energy conservation policies promoting new energy-saving physical capital would be predicted to have the desired effect. However, if they are complements, then rising energy prices would adversely effect capital formation and, hence, such policies would be counterproductive. Even without an explicit energy tax, a complementary relationship is generally more concerning to economists, in particular in the present times of high energy prices.

Table 4: Price elasticities matrices at 1947

Unrestricted SUR					Local Approach			
	mode estimates					mode estimates		
	K	L	E	M	K	L	E	M
K	-0.0974	0.4461	-0.1315	-0.2172	-0.1488	0.2897	-0.1412	0.0004
L	0.0921	-0.1774	0.0922	-0.0069	0.0598	-0.4092	0.0888	0.2606
E	-0.1579	0.5362	-0.6167	0.2384	-0.1696	0.5163	-0.6659	0.3192
M	-0.0168	-0.0026	0.0154	0.0040	0.0000	0.0978	0.0206	-0.1184
	mean estimates					mean estimates		
K	-0.0978	0.4384	-0.1320	-0.2086	-0.1485	0.2896	-0.1412	0.0002
L	0.0905	-0.1909	0.0926	0.0078	0.0598	-0.4110	0.0888	0.2624
E	-0.1585	0.5384	-0.6136	0.2337	-0.1696	0.5160	-0.6655	0.3191
M	-0.0162	0.0029	0.0151	-0.0018	0.0000	0.0985	0.0206	-0.1191
	std. of posts for distributions					std. of posts for distributions		
K	0.0625	0.1576	0.0422	0.1998	0.0492	0.1228	0.0415	0.1422
L	0.0325	0.2355	0.0402	0.2393	0.0253	0.1751	0.0392	0.1689
E	0.0507	0.2338	0.1284	0.2067	0.0498	0.2280	0.1203	0.2006
M	0.0155	0.0898	0.0133	0.0994	0.0110	0.0634	0.0129	0.0665
Global concavity					Regional approach on cube1			
	mode estimates					mode estimates		
	K	L	E	M	K	L	E	M
K	-0.1369	0.0685	0.0145	0.0539	-0.1488	0.2897	-0.1412	0.0004
L	0.0141	-0.3349	0.0569	0.2638	0.0598	-0.4092	0.0888	0.2606
E	0.0174	0.3307	-0.5425	0.1943	-0.1696	0.5163	-0.6659	0.3192
M	0.0042	0.0990	0.0125	-0.1157	0.0000	0.0978	0.0206	-0.1184
	mean estimates					mean estimates		
K	-0.1364	0.0682	0.0145	0.0537	-0.1672	0.2300	-0.1312	0.0683
L	0.0141	-0.3334	0.0570	0.2624	0.0475	-0.3898	0.0831	0.2592
E	0.0174	0.3311	-0.5431	0.1945	-0.1575	0.4830	-0.6566	0.3310
M	0.0042	0.0984	0.0126	-0.1151	0.0053	0.0973	0.0214	-0.1239
	std. of posts for distributions					std. of posts for distributions		
K	0.0502	0.0437	0.0141	0.0417	0.0468	0.1046	0.0389	0.1239
L	0.0090	0.1712	0.0289	0.1676	0.0216	0.1721	0.0339	0.1720
E	0.0169	0.1681	0.1307	0.1208	0.0467	0.1971	0.1109	0.1864
M	0.0032	0.0629	0.0078	0.0638	0.0096	0.0645	0.0120	0.0691
Regional approach on cube2					String approach			
	mode estimates					mode estimates		
	K	L	E	M	K	L	E	M
K	-0.1605	0.1938	-0.0585	0.0251	-0.1488	0.2897	-0.1412	0.0004
L	0.0400	-0.3584	0.0915	0.2268	0.0598	-0.4092	0.0888	0.2606
E	-0.0703	0.5321	-0.6585	0.1966	-0.1696	0.5163	-0.6659	0.3192
M	0.0019	0.0851	0.0127	-0.0997	0.0000	0.0978	0.0206	-0.1184
	mean estimates					mean estimates		
K	-0.1613	0.1237	-0.0398	0.0774	-0.1474	0.2900	-0.1429	0.0004
L	0.0255	-0.3661	0.0691	0.2715	0.0599	-0.4215	0.0886	0.2730
E	-0.0477	0.4015	-0.5641	0.2103	-0.1716	0.5153	-0.6714	0.3277
M	0.0060	0.1019	0.0136	-0.1214	0.0000	0.1024	0.0211	-0.1236
	std. of posts for distributions					std. of posts for distributions		
K	0.0483	0.0729	0.0228	0.0781	0.0475	0.1246	0.0406	0.1443
L	0.0151	0.1599	0.0330	0.1575	0.0257	0.1748	0.0397	0.1725
E	0.0274	0.1921	0.1100	0.1610	0.0487	0.2308	0.1204	0.1992
M	0.0061	0.0591	0.0104	0.0616	0.0112	0.0647	0.0129	0.0698

Fig. 3: Posterior distributions of the elasticity of substitution between energy and capital



In Fig. 3 we present the results of the cross price elasticity of energy with respect to capital e_{EK} under four different levels of constraints. The global concavity approach produces the far most right distribution having a mode value at $e_{EK} = 0.0145$ (see Table 4) suggesting that capital and energy are weak substitutes. The entirely data driven ‘unconstrained’ approach would suggest that capital and energy are complements. Since the unconstrained estimate however violates duality theory, one has to be very careful with such a conclusion. The regularity preserving string approach produces the distribution only slightly to the left of the unconstrained approach with the mode at $e_{EK} = -0.1412$. Instead, the cube approach produces a distribution, which is likely to be biased towards zero.

Table 5: Estimated changes in capital stock in millions of U.S. dollars in manufacturing sector due to 10% increase in energy price

	year	Global concavity approach			String approach		
		(lower 5% bound)	point estimate	(upper 95% bound)	(lower 5% bound)	point estimate	(upper 95% bound)
elasticity e_{EK}		(0.00102)	0.01450	(0.04007)	(-0.20928)	-0.14120	(-0.07555)
change in capital stock	1947	(1.0)	13.5	(37.3)	(-194.8)	-131.5	(-70.3)
	2001	(45.6)	646.0	(1785.2)	(-9323.9)	-6290.7	(-3365.7)

In the first row calculations are based on the year 1947. This represents our results comparably to the tables and figures provided in previous studies (such as Berndt and Wood (1979), Diewert and Wales (1984), Barnett et al. (1991) and Terrell (1996)). Numbers in parenthesis provide the 90% coverage probability intervals of respective changes and elasticities. Intervals are computed by using 100,000 MHARA simulator outcomes.

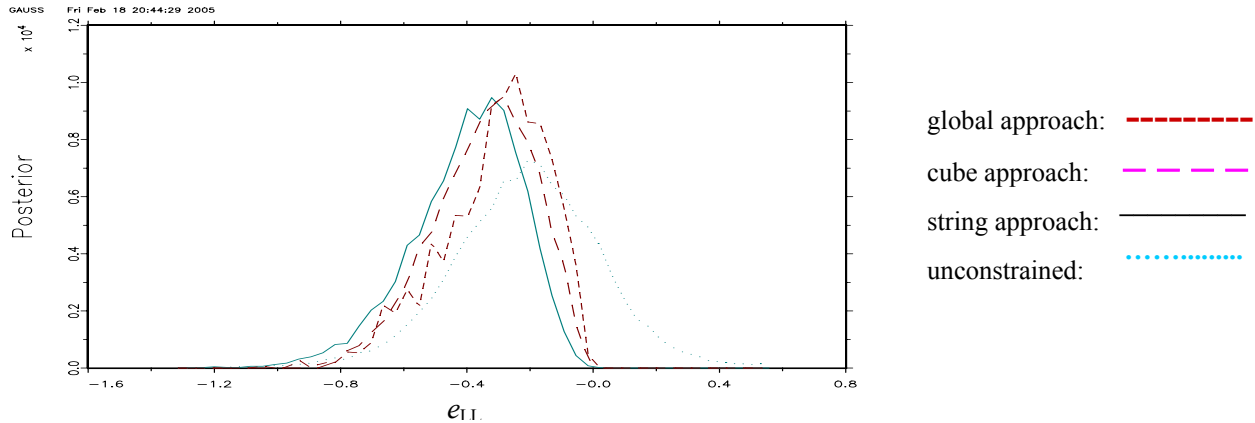
Table 5 illustrates the enormous consequences of using different estimators. Results using the cross price elasticity estimate of the standard global concavity approach imply that a 10% increase in

energy price leads in the U.S. manufacturing sector to a 14 million U.S. dollars *increase* in capital formation. Instead using our preferred MHARA string approach predicts a significant *decrease* of the capital stock by about 132 million dollars. The second row uses the most recent 2001 data provided by the Bureau of Labor Statistics. Due to the use of the different estimators, the same calculations lead to an absolute change in capital stock of about 7 billion U.S. dollars ($= 6.3 + 0.6$). Given that in the year 2001 the total value of the capital stock in the manufacturing sector amounts to 446 billion dollars, the change of 7 billion solely due to the use of different estimators is alarming. Finally note that the changes between the global approach and the string approach are statistically very significant.¹⁴ This demonstrates that when assessing the costs and benefits of an environmental tax, one must be very careful about the estimator choice because the differences in significance and the substitution patterns could lead to dramatically different policy recommendations. In particular here the global approach leads to very misleading results.

Three more observations are notable. Firstly, comparing the spreads of the distributions a robust pattern arises. The larger ψ , the smaller is the sample variance of the posterior distribution. Although one might be attracted by an estimator with a small variance, choosing the estimator on this basis would be very misleading. Here as can be seen from table 3, the variance of the estimator is rather inversely related to the model fit statistics. For a simple proof, that expanding the regularity imposing set, *ceteris paribus*, decreases the supremum of any statistical criterion functions measuring the model fit, see WHM, lemma 1.

¹⁴ This can also be seen from Fig. 3 by noting that the joint overlap of the respected estimated distributions contains little probability mass only, the 5 and 95 percentiles are displayed in Table 4.

Fig. 4: Posterior distributions of the own price elasticity of demand for labor



Secondly, from Fig. 3, an apparent pattern suggests that the starker the restrictions, the greater is the difference in the relative positions between the unrestricted approach and the restricted approaches. One therefore could conclude that, the unrestricted approach might be a good ‘approximation’, since it seems to be close to the string approach. Although for many parameters this correlation seems to be holding, there exist some important exceptions destroying this analogy. For example, the own price elasticity of material of the unconstrained approach is positive (see table 4). In contrast all regular distributions are truncated at the value of 0 assigning zero probability to the positive orthant. As another example consider Fig. 4, which displays the posterior distributions of e_{LL} . Even though the unconstrained mode of e_{LL} lies in the regular nonpositive orthant, about 20% of its probability mass falls into the irregular positive area. The posterior with the strongest left inclination is produced by the string approach, followed by the cube approach and the global approach. This is a completely different ordering of the relative positions of the distributions as for e_{EK} .

Thirdly, the distribution of the elasticity estimators are not normal; in some instances it would be misleading to use the standard errors in table 4 to construct symmetric confidence intervals. One would be better advised to use the percentiles of the posterior distributions.

6.3. Does the KLEM data set support duality theory?

6.3.1 Testing Duality Theory using a regionally regular estimate

For the KLEM data set, section 6.1 demonstrated that regularity imposing estimators are required to make the empirical model consistent with the assumed underlying economic theory. Forcing the data into such theoretical relationships should raise serious concerns whether the empirical evidence supports the duality theory, or if the economic theory should be rejected. In order to investigate this issue, a simple hypothesis test can be carried out. We test the unrestricted estimate \mathbf{b}^u (table 3, column 1) against the hypothesis of ‘duality theory’. The null of ‘duality theory’ hypothesis is represented by the regionally regular estimate \mathbf{b}^r of the string approach (since \mathbf{b}^r , (column 8) satisfies all shape restrictions, the model $c(z, \mathbf{b}^r)$ is consistent with duality theory). F, Wald and Likelihood Ratio tests (with Bartlett correction) are carried out. The duality theory hypothesis is not rejected by any test at the 5% and 10% significance levels. Similarly, using the uninformative Bayes factor of 1, the posterior odd ratio in favor of the well-behaved model is 0.874 (Zellner, 1971).¹⁵ This leads to the conclusion that the KLEM data in deed might have been generated by an underlying cost function that is consistent with duality theory.

6.3.2 Over-rejection problem using standard estimates

As indicated in section 3, the availability of regularity preserving techniques leads to improved hypothesis testing. This point is now demonstrated with the KLEM data. Suppose no regional regular MCMC estimates were available. Then, to carry out a duality theory hypothesis test, the only alternative would be to use the globally regular parameter \mathbf{b}^{gr} (column 2b) as representing the null (since all other parameter estimates violate the implied shape conditions of the duality theory, none of these other estimates could represent the null). Testing \mathbf{b}^u against \mathbf{b}^{gr} however leads to over-rejection. This is due to the fact that \mathbf{b}^{gr} is a member of the much smaller subset $\Theta^R | \mathcal{R}_+^{2+N} \subset \Theta^R | \Psi$. To show this fact for the KLEM data, we repeated the above procedure (from section 6.3.1) by testing \mathbf{b}^u against \mathbf{b}^{gr} . In stark contrast to the above findings, here the F, Wald and Likelihood ratio test results erroneously would lead to the conclusion that the duality theory hypothesis ought to be rejected. With

¹⁵ A common alternative is to report the percentage of data points where regularity violations occur. This percentage, however, should not be interpreted as a test statistic. For example, despite the fact that the hypothesis tests cannot reject the regular model, our unconstrained estimate violates regularity at 100% of the data points.

only 0.189 in favor of the well behaved model, the Bayesian posterior odd gives the same result.

In conclusion, from 5.3.1 the U.S. data do indeed seem to support the duality theory hypothesis. Using standard econometric methods in 5.3.2 creates an unfortunate divide between the empirical model and the underlying economic theory. Using instead our regularity preserving MHARA estimator closes this gap.

7. Conclusion

Frequently, economic theory places shape restrictions on functional relationships between economic variables used to model technology or economic behavior. Well known examples are curvature and monotonicity restrictions that apply to indirect utility, expenditure, production, profit, and cost functions. Unfortunately, when using flexible functional forms, estimated functions frequently violate these regularity conditions. Clearly, such a gap between economic theory and the empirical model is problematic for the interpretability of the results, and especially worrisome if one wishes to derive forecasts or policy recommendations. In view of both, the need to produce theoretically consistent models and the empirical difficulties in implementation, Diewert and Wales 1987 observed: *One of the most vexing problems applied economists have encountered is that theoretical curvature conditions that are implied by economic theory are frequently not satisfied by the estimated cost, profit or indirect utility function.*

This paper investigates estimators that might be able to close this wedge between the empirical model and the underlying economic theory. We extend and improve upon currently available estimation methods for maintaining shape conditions by imposing restrictions on a connected subset of the domain associated with the function being estimated.

Our technique is illustrated by investigating elasticities for the US demand system of the manufacturing sector. We apply a series of alternative shape imposing estimators. In comparison to these, it is shown that our technique maintains the flexibility property to the greatest possible extent, improves the goodness of fit measures and it is computationally faster. With this method we successfully produce an empirical demand system that is entirely consistent with the underlying economic theory. As a demonstration of its empirical relevance, we produce various sets of elasticity estimates and their respective distributions and

interpret differences in the light of the new methodological advances.

Finally we motivate a testing procedure that checks the plausibility of the assumed economic theory. We show that standard econometric applications could erroneously reject the hypothesis that the observed U.S. input data emerge from a true data generation process consistent with duality theory. Instead using a regional regular estimate overwhelmingly supports the assumption that the KLEM data can be modelled within a system for which all regularity conditions hold.

The benefits of the technique described in this paper can be extended to other areas of economics. For example the method could be applied to the estimation of producer supply or consumer demand systems, which also underlie multiple shape conditions implied by economic theory. Equivalently many functional relations in game theoretic models exhibit curvature, quasi-convexity or monotonicity restrictions. Also, it would also be interesting to compare these estimation results with the currently developing new techniques in nonparametric estimation that attempt to impose shape restrictions. This is to be explored in future research. We hope that the methods demonstrated in this paper promote tractability and facilitate the analysis of empirical models for which consistency with an underlying economic theory is required.

References

- Aït-Sahalia, Y. and J. Duarte (2003): Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116, pp.9 – 47.
- Andrews, D. (2001): Testing When a Parameter Is on the Boundary of the Maintained Hypothesis. *Econometrica*, 69, 683–734.
- Andrews, D. (1999): Estimation When a Parameter Is on a Boundary. *Econometrica*, 67, 1341–1383.
- Apostolakis, B.E. (1990): Energy-Capital Substitutability / Complementarity: The Dichotomy. *Energy Economics*, 1:48-58.
- Barnett, W.A., J. Geweke, and M. Wolfe (1991): Semiparametric Bayesian Estimation of the Asymptotically Ideal Production Model. *Journal of Econometrics*, 49, No. 1/2, pp.5-50.
- Antras, P. (2004): *Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution. Contributions to Macroeconomics*, Vol. 4, No. 1.
- Barnett, W.A. (2002): Tastes and technology: curvature is not sufficient for regularity. *Journal of Econometrics*, 108, 199-202.
- Barnett, W.A. and J.M. Binner (2004): *Functional Structure and Approximations in Econometrics*. Elsevier.
- Barnett, W.A. and M. Pasupathy (2003): Regularity of the Generalized Quadratic Production Model: A Counterexample. *Econometric Reviews*, 22-2, pp. 135-154.

- Berndt, E.R. and M.S. Khaled (1979): Parametric Productivity Measurement and Choice among Flexible Functional Forms, *Journal of Political Economy*, **87**, 1220-1245.
- Berndt, E.R. and D.O. Wood (1975): Technology, Prices and the Derived Demand for Energy, *The Review of Economics and Statistics*, **57**, 259-268.
- Blundell, R (2004): Presidential address of the EEA/ESEM (*European Economic Association and Econometric Society European Meeting*) conference in Madrid, August, 20, 2004.
- Chambers (1988): Applied Production Analysis, The Dual Analysis. Cambridge University Press.
- Chichilnisky, G. and G.Heal (1983): Energy Capital Substitution: a General Equilibrium Analysis. IIASA, Laxenburg, Austria.
- Dasgupta, P.S. and G.M. Heal (1979): Economic Theory and Exhaustible Resources. Cambridge University Press.
- Diewert W.E. (2004): Preface. In: Barnett, W.A. and J.M. Binner (2004): Functional Structure and Approximations in Econometrics. Elsevier.
- Diewert, W. E., and Wales, T. J. (1987): Flexible Functional Forms and Global Curvature Conditions. *Econometrica* **55**, 43-68.
- Edwards, D. and D. Terrell (2004): Does Theory Matter: Assessing the Impact of Monotonicity and Concavity Constraints on Forecasting Accuracy. Midwest Econometrics Group, *14th Annual Meeting, Northwestern University, Evanston, IL*.
- Evans, D. S., and J. J. Heckman. (1984): A Test for Subadditivity of the Cost Function with an Application to the Bell System. *American Economic Review* **74**, 615–623.
- Evans, D. S., and J. J. Heckman. (1986): A Test for Subadditivity of the Cost Function with an Application to the Bell System: Erratum. *American Economic Review* **76**, 856–858
- Gallant, A. R. and G.H. Golub (1984): Imposing Curvature Restrictions on Flexible Functional Forms. *Journal of Econometrics*, **26**, pp.295-322.
- Friesen, J. (1992): Testing Dynamic Specification of Factor Demand Equations for U.S. Manufacturing. *Review of Economics and Statistics*, Vol. **74**,2,240-50.
- Griffiths, W.E., C.J. O'Donnell and A. Tan-Cruz (2000): Imposing Regularity Conditions on a System of Cost and Factor Share Equations. *Australian Journal of Agricultural and Resource Economics*, **44**, pp.107-127.
- Geweke, J. (1986): Exact Inference in the Inequality Constrained Normal Linear Regression Model. *Journal of Applied Econometrics*, vol. 1, issue 2, pp. 127-41.
- Hsieh, C. (2000): Measuring Biased Technological Change,.Princeton University, mimeo.
- Koebel, B., M. Falk and F. Lasney (2003): Imposing and Testing Curvature Conditions on a Box-Cox Cost Function. *Journal of Business and Economic Statistics*, **21**(2), pp.319-35.
- Kumbhakar, S.C. and E.G. Tsionas (*forthcoming*): Measuring Technical and Allocative Inefficiency in the Translog Cost System: A Bayesian Approach, *Journal of Econometrics*.
- Lau, L.J. (1978): Testing and Imposing Monotonicity, Convexity, and Quasi-Convexity Constraints. In: Production Economics: A Dual Approach to Theory and Applications (volume 1), eds. Fuss, M and D. McFadden, Amsterdam: North-Holland, 1978, pp.409-453.
- Lau, L.J. (1986): Functional Forms in Econometric Model Building. Chapter 26 in Handbook of Econometrics, Vol. 3, Ed. Griliches, Z. and Intriligator, M.D., Elsevier Science, Amsterdam, North Holland, pp.1515-1566.
- Mankiw, G.N., D. Romer, and D. Weil (1992): A Contribution to the Empirics of Economic Growth,.Quarterly Journal of Economics, **107**:2, pp. 407-437.
- Matzkin, R.L. (1994): Restrictions of Economic Theory in Nonparametric Methods. Handbook of

Econometrics, Volume 4, chapter 42, Ed. R.F. Engle and D.L. McFadden, pp. 2524-58.

Moschini, G (1999): Imposing Local Curvature Conditions in Flexible Demand System, *Journal of Business & Economic Statistics*, 17:487-490.

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

Salvanes, K.G. and S. Tjøtta (1998): A Note on the Importance of Testing for Regularities for Estimated Flexible Functional Forms. *Journal of Productivity Analysis*, 9, 133-143.

Sang V Nguyen & Mary L Streitwieser (1997): Capital-Energy Substitution Revisited: New Evidence From Micro Data. *Economic Studies*, 97-4.

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

Tripathi, G. (2000): Local Semiparametric Efficiency Bounds under Shape Restrictions. *Econometric Theory*, 16, pp. 729–739.

Wolff, H., T. Heckelei, and R.C. Mittelhammer (2009): Imposing Curvature and Monotonicity on Flexible Functional Forms: An Efficient Regional Approach. Working Paper.

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