Testing Simulation and Structural Models
with Applications to Energy Demand

by

Hendrik Wolff

Diplom (University of Bonn, Germany) 2000
M.S. (University of California, Berkeley) 2004

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Agricultural and Resource Economics
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Maximilian Auffhammer, Co-chair
Professor Michael Hanemann, Co-chair
Professor David Zilberman
Professor Enrico Moretti

Spring 2007

This dissertation of Hendrik Wolff is approved:

Co-chair ______________________________________ Date __________

Co-chair ______________________________________ Date __________

______________________________________ Date __________

______________________________________ Date __________

University of California, Berkeley

Spring 2007

Testing Simulation and Structural Models with Applications to Energy Demand

by Hendrik Wolff

Abstract

Testing Simulation and Structural Models

with Applications to Energy Demand

by

Hendrik Wolff

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Maximilian Auffhammer, Co-chair

Professor Michael Hanemann, Co-chair

This dissertation deals with energy demand and consists of two parts. Part one proposes a unified econometric framework for modeling energy demand and examples illustrate the benefits of the technique by estimating the elasticity of substitution between energy and capital. Part two assesses the energy conservation policy of Daylight Saving Time and empirically tests the performance of electricity simulation.

In particular, the chapter “Imposing Monotonicity and Curvature on Flexible Functional Forms” proposes an estimator for inference using structural models derived from economic theory. This is motivated by the fact that in many areas of economic analysis theory restricts the shape as well as other characteristics of functions used to represent economic constructs. Specific contributions are (a) to increase the computational speed and tractability of imposing regularity conditions, (b) to provide regularity preserving point estimates, (c) to avoid biases existent in previous applications, and (d) to illustrate the benefits of our approach via numerical simulation

results. The chapter "Can We Close the Gap between the Empirical Model and Economic Theory" discusses the more fundamental question of whether the imposition of a particular theory to a dataset is justified. I propose a hypothesis test to examine whether the estimated empirical model is consistent with the assumed economic theory. Although the proposed methodology could be applied to a wide set of economic models, this is particularly relevant for estimating policy parameters that affect energy markets. This is demonstrated by estimating the Slutsky matrix and the elasticity of substitution between energy and capital, which are crucial parameters used in computable general equilibrium models analyzing energy demand and the impacts of environmental regulations. Using the Berndt and Wood dataset, I find that capital and energy are complements and that the data are significantly consistent with duality theory. Both results would not necessarily be achieved using standard econometric methods.

The final chapter "Daylight Time and Energy" uses a quasi-experiment to evaluate a popular energy conservation policy: we challenge the conventional wisdom that extending Daylight Saving Time (DST) reduces energy demand. Using detailed panel data on half-hourly electricity consumption, prices, and weather conditions from four Australian states we employ a novel 'triple-difference' technique to test the electricity-saving hypothesis. We show that the extension failed to reduce electricity demand and instead increased electricity prices. We also apply the most sophisticated electricity simulation model available in the literature to the Australian data. We find that prior simulation models significantly overstate electricity savings. Our results suggest that extending DST will fail as an instrument to save energy resources.

______________________________________ Date __________

Co-Chair Professor Maximilian Auffhammer

______________________________________ Date __________

Co-Chair Professor Michael Hanemann

To Ada

Contents

List of Figures

List of Tables

Acknowledgements

My time at Berkeley working on my Ph.D. has been intense and, at the same time, exciting and enlightening. I have greatly enjoyed it.

This experience would have been impossible without the help of two very important advisors, Maximilian Auffhammer and Michael Hanemann, who both lead by example.

Michael, thank you for your encouraging way of guiding me to a deeper understanding of knowledge and work. Thanks for your Blackberry correspondence (we both were out of town quite a bit), and after our personal conversations I always felt freshly inspired. In fact, I can’t imagine requesting a better teacher in Environmental Economics.

Max, thank you for your friendship and the always warm, motivating, and stimulating way you interacted on our research projects. I felt that I could always talk to you, from just bouncing economic ideas off you to asking the most technical questions on model equations. This has been an invaluable asset to me and it strongly improved my effectiveness and joy at work. I will strive, as a professor, to give my students the same motivation as you gave to me.

I also give a special thanks to Guido Imbens, Jeffey LaFrance, Enrico Moretti, and David Zilberman, for many fruitful discussions and for offering me their advice on a continued basis, as well as their assistance in developing my ideas. I am deeply indebted for their goodwill and many insights.

I am very happy to have had really terrific fellow students. Especially, working with Ryan Kellogg on the Daylight Savings Time and Energy project not only broadened my understanding of research, but also made the work much more enjoyable. This produced one joint paper that is now part of this thesis. Also, it led to really fun presentations at numerous conferences and seminars and generated seemingly endless media coverage around the world. That has been a new and exciting experience.

I cooperated with Thomas Heckelei and Ron Mittelhammer on my work on shape restrictions. These long-time friends in fact had a considerable impact on my decision to pursue a doctorate in the United States and so I thank them for their invaluable encouragement at the beginning and their collaboration throughout.

I am grateful to everyone that has read parts of the manuscript, or discussed their ideas with me, especially Michael Anderson, Severin Borenstein, Jennifer Brown, Kenneth Chay, Ann Harrison, Jeffrey Perloff, Muzhe Yang, Arnold Zellner and the late Jean Lanjouw.

I thank Alison McDonald and Lesley Rowland from Australia for helping understand the electricity and weather data, and Adrienne Kandel for conversations regarding the details of the California electricity simulation model. Additional thanks are due Ernst Berndt, for providing the manufacturing data.

Thanks to all friends and colleagues for providing such a great working and leisure atmosphere. Many contributed with their unique personalities to make life in Berkeley such a fantastic experience. Especially, I want to mention Arturo, Celine, David, Damon, Felipe, Kostas, Kristin, Muzhe, Ralf and of course the ‘sexy men’ Maoyong and Ricardo. Special thanks goes to my office mate Emma for putting up with my working habits and to Rob for his patience at the bench press.

None of this would have been possible without the lasting support of my parents. Throughout, they encouraged me, never tired of asking when I would finish, and always supported me with their hearty best wishes.

Finally and most importantly, to Anne and Ada, my deepest thanks for supporting me with your love, and your understanding for my being far away from home.

Chapter 1

Introduction

With rising energy prices and pressing environmental concerns research on energy issues has become increasingly important. In particular, the growing worldwide demand for energy asks for new approaches to the analysis of energy policies and their effectiveness in terms of market outcomes, energy conservation and environmental issues. This dissertation *Testing Simulation and Structural Models with Applications to Energy Demand* contributes to these efforts and consists of three main chapters.

The second chapter proposes a unified framework for estimation of structural models derived from economic theory.[1] This work is motivated by the fact that in many areas of economic analysis, economic theory restricts the shape as well as other characteristics of functions used to represent economic constructs. Obvious examples are the monotonicity and curvature conditions that apply to utility, profit, and cost functions. Commonly, these regularity conditions are imposed either locally or globally. Here we extend and improve upon currently available estimation methods for imposing regularity conditions by imposing regularity on a connected subset of the regressor space. This method offers important advantages over the local approach by imposing theoretical consistency not only locally, at a given evaluation point but also within the whole empirically relevant region of the domain associated with the function being estimated. The method also provides benefits relative to the global

[1] With Thomas Heckelei and Ron Mittelhammer.

approach, through higher flexibility, which generally leads to a better model fit to the sample data compared to the global imposition of regularity. Specific contributions of the second chapter are (a) to increase the computational speed and tractability of imposing regularity conditions in estimation, (b) to provide regularity preserving point estimates, (c) to avoid biases existent in previous applications, and (d) to illustrate the benefits of the regional approach via numerical simulation results.

The third chapter discusses the more fundamental question of whether the imposition of a particular theory to a dataset is justified. I propose a hypothesis test to examine whether the estimated empirical model is consistent with the assumed economic theory. The test relies on the following principle: Behavior and/or assumptions on technology manifest 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 behavioral 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. A motivating application exemplifies the use of the techniques developed in the first two chapters. Climate change concerns drive many countries to debate over imposing a tax on energy use intended to reduce carbon dioxide emissions. To quantitatively assess the costs and benefits of such a policy, an analyst requires two pieces of 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. As another example the elasticity of substitution between capital and energy is considered. This is an important policy parameter as knowledge over the effect on the energy market as a response to a change in interest rate is a re-occurring question. The advantages of the proposed estimation and testing methods are hence illustrated using the Berndt and Wood industry data. I show that this dataset is consistent with the assumption that firms minimize costs using standard microeconomic theory and I demonstrate that energy and capital are complements—a result that has been much debated. Both findings would not necessarily be obtained when employing standard econometric methods to this dataset and hence has potentially important policy implications.

The fourth chapter uses a quasi-experiment and challenges the conventional wisdom that extending Daylight Saving Time reduces energy demand. The third chapter *Daylight Time and Energy*[2], analyzes the effect of Daylight Saving Time (DST) on electricity demand. This work is motivated by the fact that the U.S.

[2] With Ryan Kellogg, fellow student at UC Berkeley.

beginning in 2007 will lengthen DST by one month, with the stated goal of reducing electricity consumption by 1%. Similar proposals are under consideration in Australia, Great Britain and Japan where the reduction of greenhouse gas emissions is the primary policy objective; California has even petitioned for year-round DST, projecting savings of up to $1.3 billion annually. We question the stated savings from prior DST studies, since they largely rely on simulation models rather than empirical evidence. Our research exploits a natural experiment, in which parts of Australia extended DST by two months to facilitate the Sydney Olympic Games in 2000. Because the Olympics can directly affect the electricity demand we focus on Victoria, which did not host Olympic events, as the treated state and use its neighbor state South Australia, which did not extend DST, as the control. Using detailed panel data on half-hourly electricity consumption, prices, and weather conditions from four Australian states we employ a novel 'triple-difference' technique to test the electricity-saving hypothesis. We show that the extension failed to reduce electricity demand and instead increased electricity prices. We also scrutinize prior DST studies and apply the most sophisticated electricity simulation model available in the literature to the Australian data. We find that all prior models significantly overstate electricity savings. Our results suggest that DST will fail as an instrument to save energy resources.

Chapter 2

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

2.1 Motivation and Literature Review

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, Geweke and Wolfe, 1991, Koop, Osiewalski and Steel, 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 nonparametric estimation techniques that account for shape restrictions (originally proposed by Hildreths, 1954) have garnered increasing attention in the literature (Matzkin, 1994, Tripathi 2000, Aït-Sahalia and Duarte, 2003). 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. Given these potential disadvantages, 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. Violations can lead to ambiguous forecasts and errant conclusions about economic behavior. Concerns related to the imposition of regularity conditions is as old as the literature on flexible functional forms and represents '*one of the most vexing problems applied economists have encountered*' Diewert and Wales (1987).

In this chapter we propose and illustrate a Bayesian estimation procedure for imposing regularity conditions via nonlinear inequality constraints. The conditions are imposed on a connected[3] 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

[3] 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 $\mathbf{U} \cup \mathbf{V} = \mathbf{X}$.

the *regional*, *local* and *global* approach, we begin by discussing how previous methods handled the imposition of regularity.

2.1.1 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[4] flexible functional forms, however, such restrictions come at the cost of limiting the flexibility of the functional form with regard to representing other economic relationships. For example, under the imposition of global concavity, the Generalized Leontief cost function does not allow for complementary relationships among inputs.

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

[4] 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.

local flexibility), which is a fact that does not seem to be generally appreciated in the literature on globally flexible functional forms.[5]

2.1.2 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,6 that can also be rife with statistical testing/verification problems. In 1984, GALLANT and GOLUB proposed an inequality constrained optimization program to impose regularity conditions locally at each observed regressor value. Compared with the global approach, this method generally increases the fit of the model to the data. However, two problems remain: (a) the procedure becomes numerically difficult for large sample sizes and/or complicated constraints and (b) it is possible that the estimated form is irregular at points other than

---

[5] 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.

[6] Examples of functional forms investigated are the Minflex Laurent (Barnett 1985), Extended Generalized Cobb Douglas (Magnus, 1979), Symmetric Generalized McFadden and Symmetric Generalized Barnett (Diewert and Wales 1987). Furthermore see the cited literature in Barnett, Geweke and Wolfe (1991:p.10) and more recently Terrell (1995, 1996), Ivaldi et al. (1996), Fleissig, Kastens and Terrell (1997, 2000), Jensen (1997), Ryan and Wales (1998), Fischer, Fleissig and Serletis (2001) for studies evaluating these mentioned and other competing forms. We recommend Barnett, Geweke and Wolfe (1991: pp.3-15) for an extensive and insightful review on the various developments, trials and errors in the history of using flexible functional forms.

the sample observations. Hence, more general methods of imposing the regularity conditions are desirable and those which appear to be the most promising are summarized below in section 2.1.3.

2.1.3 Towards regional regularity

In order to circumvent the problem of the estimated form being irregular at points other than the sample observations, GALLANT and GOLUB discussed the possibility of imposing regularity conditions on a predefined regular region $\boldsymbol{\psi}$ 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 did not demonstrate the tractability of this approach and it seems that empirical implementation can be formidable with the currently available optimization tools.

It was not until 1996 that TERRELL advanced ideas relating to the empirical application of regional regularity. Instead of explicitly using a constrained optimization algorithm he decomposed the problem into a series of steps: First, a convex set $\boldsymbol{\psi}$ 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 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, O'Donnell and Tan Cruz (2000:p.116) suggested using a Metropolis-Hastings Accept-Reject Algorithm (subsequently denoted as MHARA). Compared to the Gibbs algorithm, MHARA may increase the probability that sampled parameter vectors are regular, and therefore may be faster than Gibbs sampling. However, the related literature on MHARA[7] did not pursue the regional approach further, but rather continued to impose local regularity without proving the theoretical consistency on the domain of interest.

[7] Literature on applications of MHARA include Koop, Osiewalski and Steel (1994), O'Donnell, Shumway and Ball (1999), Griffiths, O'Donnell and Tan Cruz (2000), Griffiths (2003), Chua, Griffiths and O'Donnell (2001), Cuesta et al. (2001), Kleit and Terrell (2001), O'Donnell, Rambaldi and Doran (2001) and O'Donnell and Coelli (2003).

2.1.4 Objectives and organization of Chapter 2

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 Bayesian approach with the MHARA. This defines an alternative methodology that substantially mitigates previous difficulties and inconsistencies in applying the regional regularity concept. New features of our proposed method include:

1. a set of sufficient conditions for which regularity is guaranteed at 'any' point in $\boldsymbol{\psi}$ (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).

3. demonstrating that the commonly used MHARA sampling technique suffers from an upward bias of posterior density values in the neighborhood of the truncation boundary. We provide a simple bias-mitigating alternative.
4. demonstrating that the commonly used posterior mean may be inappropriate as a point estimate of model parameters due to the potential violation of regularity conditions. As an alternative, we suggest two regularity-preserving point estimates:
    - i) the posterior mode
    - ii) the parameter vector that minimizes error loss subject to regularity constraints.

The organization of the chapter is as follows: In section 2, we motivate the methodology and outline the estimation procedure in general terms. Section 3 provides a more technical description of procedures and discusses the four methodological contributions. Examples using AIM functional forms are given in section 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.

2.2 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.

2.2.1 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, $c(\mathbf{p};\beta)$, whereby $\mathbf{p} = [p_1, p_2,\ldots, p_K]^\mathsf{T} \in \pi$ are K input prices, π denotes the orthant of strictly positive prices in $\Re^K$, and $\beta \in \Theta$ is the parameter vector to be estimated. According to economic theory $c(\mathbf{p};\beta)$ must be concave and nondecreasing in $\mathbf{p}$ (Mas-Colell, Whinston and Green, 1995:p.141). The regularity conditions to be imposed on a subset $\boldsymbol{\psi}$ of the price space π can be characterized by H elementary *Inequality Constraint Functions*, $\mathbf{i}(\mathbf{p};\beta) \equiv [i_1,i_2,\ldots,i_H]$: $(\pi \times \Theta) \rightarrow \Re^H$, whereby the restrictions hold whenever, for a given β, $\mathbf{i}()$ is nonnegative for all prices in the relevant region $\boldsymbol{\psi}$,

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

For example, if $c(\mathbf{p};\beta)$ 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[8]

$$i_1 = \partial c(\mathbf{p};\beta)/\partial p_1, \qquad i_2 = \partial c(\mathbf{p};\beta)/\partial p_2,$$

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

Note that previous *global* and *local* estimation methodologies differ in the way $\boldsymbol{\psi}$ is defined. If $\mathbf{i}(\mathbf{p};\beta) \geq \mathbf{0} \;\forall\; \mathbf{p} \in \boldsymbol{\psi}$, we say that regularity is imposed (i) *locally* if $\boldsymbol{\psi}$ consists of one or more singular disconnected points in π, (ii) *globally* if $\boldsymbol{\psi} = \pi$, and

[8] Note that nonnegativity of i_1 and i_2 imposes monotonicity. Nonnegativity of i_3 and i_5 imposes negative semi-definiteness on the Hessian $\partial^2 f(\mathbf{p};\beta)/\partial\mathbf{p}\partial\mathbf{p}'$. Since by linear homogeneity of $f(\cdot)$ the Hessian has rank $K - 1$, it is not necessary to generate an additional inequality constraint function to sign the K^{th} principal minor.

(iii) *regionally* if ψ is some connected subset of π. 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 ψ.

> ***Definition 1:*** *The empirically relevant set* ψ *is a closed*[9] *and connected subset of* π *that covers the empirically relevant price region, defined as containing all sample observation* $n = 1,\ldots,N$ *as well as any price points* $c = 1,\ldots,C$ *that will be used for subsequent analyses and/or simulations based on the estimated model*.

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

Fig 1: Irregular cost function

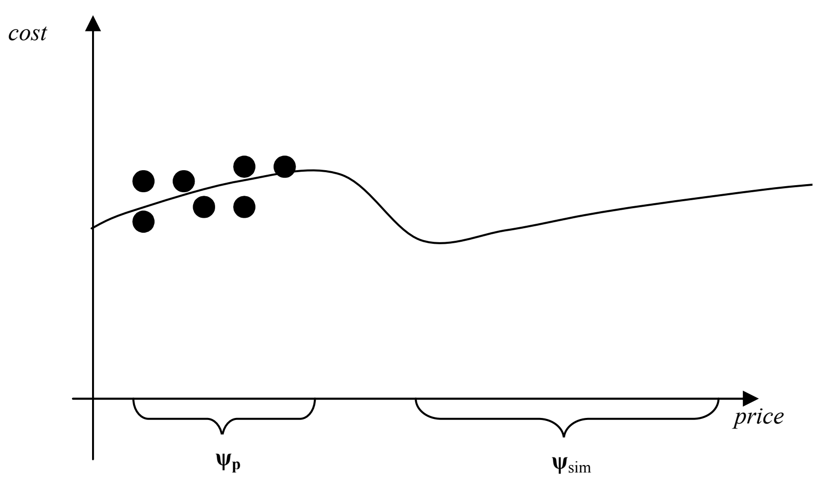


Fig. 1 depicts an example where ψ_p includes all observed data points (each dot represents an observed (cost, price) combination used for estimating the cost function), and ψ_{sim} includes the region at which inferences will be drawn for simulation purposes. However, $\psi = \psi_p \cup \psi_{sim}$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

[9] The requirement that ψ is a closed set simplifies the proofs of some later propositions, but is not necessary for any other reason.

2.2.2 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.[10] We assume that $\boldsymbol{\varepsilon}$ is an $M{\cdot}N \times 1$ unknown error vector with mean $\mathrm{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Further, $\boldsymbol{\Theta}$ is the L-dimensional parameter space, which, if the regularity conditions are to hold for all values of $\mathbf{p}$ in $\boldsymbol{\psi}$, reduces to the L-dimensional regular subset $\boldsymbol{\Theta}^{\mathrm{R}} \subset \boldsymbol{\Theta}$ defined as[11]

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

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

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

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,

---

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

[11] 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 $\boldsymbol{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}_* \cup \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_* = \boldsymbol{\Theta}$.

$p(\mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-(M+1)/2}$, and further assuming that $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$ are a priori independent, the joint prior is defined as

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

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 $\mathbf{\Theta}^{R}|\boldsymbol{\psi}$ is dependent on the choice of $\boldsymbol{\psi}$, the marginal conditional improper[12] prior on the $\boldsymbol{\beta}$ vector is specified as an indicator function

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

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 $\mathbf{\Theta}^{R}|\boldsymbol{\psi}$ 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.

[12] 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, $1\{\boldsymbol{\beta}\in\mathbf{\Theta}^{R}|\boldsymbol{\psi}\}$ is technically not a "proper" prior distribution. It is not normalized to integrate to 1, and moreover, if $\mathbf{\Theta}^{R}|\boldsymbol{\psi}$ does not have finite volume, $\int p(\boldsymbol{\beta}|\boldsymbol{\psi})d\boldsymbol{\beta} = \infty$. 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 $\mathbf{\Theta}^{R}|\boldsymbol{\psi}$.

2.2.3 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 Θ^{R}. The outcomes are then used to approximate posterior expectations via the appropriate empirical estimates, e.g. $J^{-1}\sum_{j=1}^{J} g(\mathbf{b}(j))$ for approximating $E[g(\boldsymbol{\beta})]$. The estimates converge to the true expectations as J increases.[13]

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. $\mathbf{b}^{(j)} \in \Theta^{R}|\boldsymbol{\psi}\ \forall\ j$. 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 $\boldsymbol{\psi}_{g} \subset \boldsymbol{\psi}$ which consists of possibly $Q \approx$ tens-of-thousands of equidistant distinct points.[14] 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 $\mathbf{b}^{(j)} \in \Theta^{R}|\boldsymbol{\psi}_{g}\ \forall\ j$, whereby $\Theta^{R}|\boldsymbol{\psi}_{g}$ is the *approximated* regularity posterior support, which will tend towards the

[13] See literature cited in footnote 15 for useful introductions into MCMC methods.

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

actual set $\boldsymbol{\Theta}^{R}|\boldsymbol{\psi}$ the finer the approximation grid $\boldsymbol{\psi}_g$. In order to circumvent the approximate nature of this representation, in a later subsection we identify problem conditions under which checking certain key points in $\boldsymbol{\psi}$ will guarantee overall regularity $\forall$ $\mathbf{p} \in \boldsymbol{\psi}$.

2.3 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 (step 1 to step 4), application of the MHARA (step 5 to step 11) and inferences based on the regularity posterior (step 12). In the subsections to follow, we explain the objectives of the steps that are nonstandard[15] and develop necessary technical details.

2.3.1 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 $p(\boldsymbol{\beta},\boldsymbol{\psi}) = 1\{\boldsymbol{\beta}\in\Theta^{R}|\boldsymbol{\psi}\}$: The regularity conditions (defined by economic theory) are identified (step 2), the empirical relevant region $\boldsymbol{\psi}$ is chosen by the researcher (step 3) and subsequently approximated by a grid $\boldsymbol{\psi}_g$ (step 4).

[15] ***Step 1, Step 5, Step 10*** and ***Step 11*** are not further elaborated on because their content is either obvious from the explanation given in box 1, or they are part of the conventional Metropolis-Hastings algorithm, which we assume the reader to be familiar with. In order to keep it is as uncomplicated as possible we outline the simplest way of implementing the Markov Chain. Other procedures like multiple chains and other proposal distributions are suggested in the literature. The reader is referred to Chib and Greenberg (1996), Richarson and Spiegelhalter (1996), Robert and Casella (1999) or Chen, Shao and Ibrahim (2000) for a further discussion of appropriate modifications of the Metropolis-Hastings algorithm.

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

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

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

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 $\mathbf{i}(\mathbf{p};\beta) \geq \mathbf{0}$. An illustrative example for the case of monotonicity and curvature restrictions was given in section 2.2.1.

Step 3: In contrast to defining $\boldsymbol{\psi}$ as one convex hyperrectangle (as in Gallant and Golub, 1984 and Terrell 1996), we define $\boldsymbol{\psi}$ as any connected (possibly non-convex) set. This has potential advantageous. 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* π *and let* s: $\Theta^{R}|\boldsymbol{\psi}_* \rightarrow \Re^1$ *be any scalar function.*

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

Suppose $s(\beta)$ 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 section 2) $\boldsymbol{\psi}$ should be defined as small as possible. This often results in a non-convex set $\boldsymbol{\psi}$ to which the methodology can be equally applied by decomposing $\boldsymbol{\psi}$ into *I convex* subsets $\boldsymbol{\psi}_i$ $\forall$ $i = 1,\ldots,I$, such that $\boldsymbol{\psi} = \bigcup_{i=1}^{I} \boldsymbol{\psi}_i$.[16] In many cases it turns out practical to construct $\boldsymbol{\psi}$ as $I = N+C$ line segments connecting all empirically relevant points thereby promising an increased fit of the estimated model to the data. For details, see the application in section 4.

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

---

[16] 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 section 4.2.

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

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

a) improve the computational speed of the algorithm and

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

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

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

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

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

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

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

Table 1 summarizes six cases for constructing sufficient “evaluation sets” $\boldsymbol{\psi}_h$, h=1,…H. In particular, the cases 2 and 5 are of interest since these enhance the computational speed considerably by reducing the number Q^* of necessary grid points to be checked. These evaluation points are described in the following definitions:

(1) The $K \times 1$ price vector $\mathbf{z}_h$ is one vertex of the hyperrectangle $\boldsymbol{\psi}$ (Q^*=1).[17]

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

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 h^{th} grid	Proposition
1	+					boundary $\mathbf{B}_h$	$\boldsymbol{\Theta}^R\vert\mathbf{B}_{gh} \supset \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	1a
2	+	+				one vertex $\mathbf{z}_h$	$\boldsymbol{\Theta}^R\vert\mathbf{z}_h = \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	1b
3			+			boundary $\mathbf{B}_h$	$\boldsymbol{\Theta}^R\vert\mathbf{B}_{gh} \supset \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	2a
4		+	+			side $\mathbf{S}_h$	$\boldsymbol{\Theta}^R\vert\mathbf{S}_{gh} \supset \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	2b
5		+		+		all vertices $\mathbf{Z}_h$	$\boldsymbol{\Theta}^R\vert\mathbf{Z}_h = \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	3
6					+	shield $\mathbf{S}^*$	$\boldsymbol{\Theta}^R\vert\mathbf{S}^* \supset \boldsymbol{\Theta}^R\vert\boldsymbol{\psi}$	4

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 $\boldsymbol{\psi}$. Considering the above five set definitions of $\boldsymbol{\psi}_h$, 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:* $i_h \geq 0\ \forall\ \mathbf{p} \in \boldsymbol{\psi}$ *iff* $i_h \geq 0\ \forall\ \mathbf{p} \in \boldsymbol{\psi}_h$ *(whereby $\boldsymbol{\psi}_h$ takes the form as indicated in the column ‘$\boldsymbol{\psi}_h$’).*
For case 6: *Suppose property V holds. Then for all inequality constraint functions* $\mathbf{i}^*(\cdot)$ *that impose nonnegative slope, nonpositive slope, concavity and/or convexity:* $\mathbf{i}^*(\cdot) \geq \mathbf{0}\ \forall\ \mathbf{p} \in \boldsymbol{\psi}$ *iff* $\mathbf{i}^*(\cdot) \geq \mathbf{0}\ \forall\ \mathbf{p} \in \mathbf{S}^*$.
For the proofs of these statements see section A1 of the appendix.

These small evaluation sets not only enormously increase the computational speed, but are of interest from a theoretical perspective as well, that is expressed in the following proposition:

[17] 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 $\partial i_h/\partial p_k \leq 0\ \forall\ \mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\max}$ and if $\partial i_h/\partial p_k \geq 0\ \forall\ \mathbf{p} \in \boldsymbol{\psi}$, then the k^{th} element of $\mathbf{z}$ is $p_k^{\min}$.

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

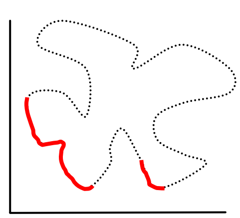
Hence in these two cases f () becomes strictly regular, instead of regular in *approximation* only as $Q \rightarrow \infty$.[18] Depending on which *properties* I-V hold, the set $\boldsymbol{\psi}_h$ can take 3 further forms of interest:

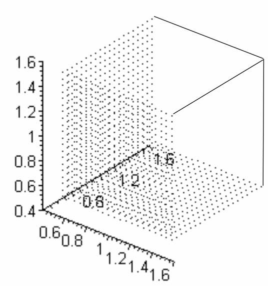
(3) $\mathbf{B}_h$ = bd($\boldsymbol{\psi}$) denotes the boundary of $\boldsymbol{\psi}$.

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

(5) $\mathbf{S}^* \subset \mathbf{B}$ is 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 $\mathbf{0} \in \pi$ (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 $\mathbf{y} \in \pi$, then $\mathbf{S}^* = \{\mathbf{p} \in \text{bd}(\boldsymbol{\psi})\text{: } \forall\ \varphi \text{ if } \varphi \in \text{bd}(\boldsymbol{\psi}) \cap l(\mathbf{0},\mathbf{p})\text{, then } ||\mathbf{p}|| \leq ||\varphi||\}$.

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





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

[18] If $Q \rightarrow \infty$, 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 $\mathbf{b} \in \boldsymbol{\Theta}^R|\boldsymbol{\psi}_g$ is such that $f(\mathbf{p};\mathbf{b})$ is almost everywhere in $\boldsymbol{\psi}$ regularity-retaining.

Two final remarks are in order: The first five cases in table 1 are independent of the "type of regularity conditions" to be imposed. Case 6, instead, is less general but applies when imposing monotonocity and curvature, (and thus suits the cost-function example in section 2 and 4). Then the shield $\mathbf{S}^*$ has to be evaluated only. Secondly, 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 $\mathbf{B}_h$ = 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 '$_g$').

2.3.2 The Metropolis-Hastings Accept-Reject algorithm and posterior bias

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

Step 6: $\mathbf{b}^{(*)}$, a candidate for the j^{th}+1 vector in the MCMC sequence $\{\mathbf{b}^{(j)}\}_{j=1}^{J+S}$, is generated by a symmetric *proposal distribution* $\delta \cdot p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$.[19] One possibility for drawing outcomes from $p(\mathbf{b}^{(*)};\mathbf{b}^{(j)})$ that accounts for *linear equality* constraints on parameters (e.g. for the symmetry condition on the Hessian $\partial^2 f(\mathbf{p};\boldsymbol{\beta})/\partial\mathbf{p}\partial\mathbf{p}'$) is to use the multivariate normal distribution $\mathbf{N}(\mathbf{b}^{(j)},\mathbf{cov}(\mathbf{b}_u))$ to generate the $L \times 1$ vector $\mathbf{b}^{(**)}$, and then to calculate

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

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

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

Step 7 and 9: Step 7 is inserted to save computing time associated with step 8 for vectors $\mathbf{b}^{(*)}$ that are already irregular at the vertices of $\boldsymbol{\psi}$. If $\mathbf{b}^{(*)}$ is identified to be irregular (either after step 7 or 9), $\mathbf{b}^{(*)}$ must be *discarded* and a new $\mathbf{b}^{(*)}$ drawn in step 6 (see the dotted arrows in box 1) using the last regular draw $\mathbf{b}^{(j)}$ as the mean of the symmetric proposal distribution $\delta{\cdot}p(\mathbf{b}^{(*)},\mathbf{b}^{(j)})$. This is repeated until $\mathbf{b}^{(*)} \in \Theta^{R}|\boldsymbol{\psi}_g$. The 'discarding' is necessary to avoid an upward bias of the regularity posterior density values in the neighborhood of the truncation boundary.[20]

To our knowledge in all previously published descriptions of the MHARA[21] it was common to *repeatedly* include the last regular $\mathbf{b}^{(j)}$ as an outcome of the simulated regularity posterior as $\mathbf{b}^{(j+1)} = \mathbf{b}^{(j)}$ until $\mathbf{b}^{(*)} \in \Theta^{R}|\boldsymbol{\psi}$. This practice, however, distorts the simulated regularity posterior in the peripheral region of $\Theta^{R}|\boldsymbol{\psi}$ close to the truncation boundary to $\Theta^{IR}|\boldsymbol{\psi}$. This is due to the fact that the probability of drawing an irregular $\mathbf{b}^{(*)}$ is higher, the closer the last regular draw $\mathbf{b}^{(j)}$ is to the frontier of $\Theta^{IR}|\boldsymbol{\psi}$.[22]

---

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

[21] Among others, the studies of O'Donnell, Shumway and Ball (1999), Griffiths, O'Donnell and Tan Cruz (2000), Griffiths (2003), Chua, Griffiths and O'Donnell (2001), and Cuesta et al. (2001), O'Donnell, Rambaldi and Doran (2001) did not account for this bias.

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

To complete step 9, if the drawn parameter vector $\mathbf{b}^{(*)}$ is regular $\forall$ $\mathbf{p} \in \boldsymbol{\psi}_g$, calculate[23]

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

Finally note that step 7 and the 'else condition' of step 9 (see the dotted arrows in box 1) *approximate* the behavior of the indicator function $1\{\boldsymbol{\beta}\in\boldsymbol{\Theta}^{R}|\boldsymbol{\psi}\}$ by subtracting $\boldsymbol{\Theta}^{IR}|\boldsymbol{\psi}_g$ (instead of $\boldsymbol{\Theta}^{IR}|\boldsymbol{\psi}$) from $\boldsymbol{\Theta}$.

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 $\boldsymbol{\psi}_{gh} = \mathbf{Z}_h$ or $\boldsymbol{\psi}_{gh} = \mathbf{z}_h$, the h^{th} evaluation of step 8 can, of course, be skipped.

2.3.3 Point estimates: inconsistency of the mean and two alternatives

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

As far as we are aware, all previous studies applying MCMC and Importance sampling to impose regularity conditions define the point estimate of $\boldsymbol{\beta}$ as the mean

---

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

E[β] of the regularity posterior.[24] However, this may result in regularity violations, as indicated in the following proposition.

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

We propose two alternative estimators that, in addition to imposing regularity (objective I), maximize a model fit measure $s(\beta)$ on $\Theta^{R}|\boldsymbol{\psi}_{g}$, 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.

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

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

of the regularity posterior as the point estimate to maximize model fit subject to the regularity conditions. To motivate $\beta^{(\text{mode})}$, note that the information contained in the normal unrestricted posterior $p_{u}(\beta|\mathbf{y}) \propto |(N\text{-}L)\boldsymbol{\Sigma}|^{-N/2}$ (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) proved that the minimization of $|\boldsymbol{\Sigma}|$

[24] These include Barnett, Geweke and Wolfe (1991), Koop, Osiewalski and Steel (1994, 1997), Terrell (1996), Terrell and Dashti (1997), O'Donnell, Shumway and Ball (1999), Griffiths, O'Donnell and Tan Cruz (2000), Chua, Griffiths and O'Donnell (2001), Kleit and Terrell (2001), Cuesta et al. (2001), Adkins, Rickman and Hameed (2002), O'Donnell, Rambaldi and Doran (2001) and O'Donnell and Coelli (2003).

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 $\boldsymbol{\beta} \in \Theta$ produces the exact same result as the maximization of $p_u(\boldsymbol{\beta}|\mathbf{y})$ over $\boldsymbol{\beta} \in \Theta$. So long as no other prior than the regularity prior is applied, we have that $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto p_u(\boldsymbol{\beta}|\mathbf{y})\cdot 1\{\boldsymbol{\beta}\in\Theta^R|\boldsymbol{\psi}\} \propto |(N\text{-}L)\Sigma|^{-N/2}$ for $\boldsymbol{\beta} \in \Theta^R|\boldsymbol{\psi}$. 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 $p_u(\mathbf{b}^{(j)}|\mathbf{y})$ $\forall\, j$ resulting from the MHARA as

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

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

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

which minimizes the posterior weighted deviation over $\boldsymbol{\beta} \in \Theta^R$, where $\|\cdot\|_\varphi$ is some vector norm[25] measuring the distance between two points within Θ^R. For example, with $\|\cdot\|_2$, the standard Euclidean norm, $\mathbf{b}^{(\text{LF}_2)} = \underset{\mathbf{b}^{(j)}}{\arg\min}\, J^{-1}\sum_{i=1}^{J}(\mathbf{b}^{(j)} - \mathbf{b}^{(i)})'(\mathbf{b}^{(j)} - \mathbf{b}^{(i)})$, minimizing the empirical-MCMC analogue to the expected squared LF subject to the regularity constraints.

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

We reemphasize that if cases 2 or 5 of table 1 apply $\forall$ h, then $\mathbf{b}^{(\mathrm{LF}_{\varphi})}$ and $\mathbf{b}^{(\mathrm{mode})}$ are members of the regular set $\mathbf{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ 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 $\mathbf{\Theta}^{\mathrm{IR}}|\boldsymbol{\psi}$, see footnote 18.

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

2.4 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).[26] In the first subsection local, global and regional regularity approaches are compared based on a specified convex set $\boldsymbol{\psi}^{\square}$. The purpose of the second subsection is to demonstrate the effects of shrinking the size of $\boldsymbol{\psi}$.

2.4.1 Experiment I - convex cube ψ

2.4.1.1 Data Generation

We now briefly describe the design of the simulations.[27] The *true* data generation process is formulated by the well-known CES cost function

$$f^{\text{CES}}(\mathbf{p};\alpha_k,\rho) = [\Sigma_{k=1}^{3} a_k^{1/(1-\rho)} \cdot p_k^{-\rho/(1-\rho)}]^{(1-\rho)/-\rho}.$$

As in Terrell, no stochastic error term is added. The derivatives result, by Shephard's Lemma, in $K = 3$ input demand functions,

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

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

---

[26] 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 $\mathbf{p}$.

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

2.4.1.2 Estimation and Evaluation

The purpose of the first experiment is to assess potential advantages of the regional approach compared to the local and global approach both in terms of model fit and the propensity for regularity violations. The normal SUR system of K = 3 input demand functions, $\hat{\mathbf{x}}_k = \partial \mathbf{f}_k^{\mathbf{AIM}(\tau)}(\mathbf{P};\hat{\boldsymbol{\beta}})/\partial p_k + \hat{\mathbf{u}}_k$ is estimated, whereby $\hat{\mathbf{u}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ represents the 64 × 1 approximation error vector to the 'true' data generation process (7), $L < N$ [28] and $\hat{\mathbf{x}}_k$ is the estimated k^{th} 64 × 1 input demand vector derived from the Asymptotically Ideal Production Model, AIM(τ), with

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

$$f^{\text{AIM}(2)} = \Sigma_{k=1}^{3} \beta_k p_k + \beta_4 p_1^{3/4} p_2^{1/4} + \beta_5 p_1^{3/4} p_3^{1/4} + \beta_6 p_1^{1/2} p_2^{1/2} + \beta_7 p_1^{1/2} p_2^{1/4} p_3^{1/4} +$$

$$\beta_8 p_1^{1/2} p_3^{1/2} + \beta_9 p_1^{1/4} p_2^{3/4} + \beta_{10} p_1^{1/4} p_2^{1/2} p_3^{1/4} + \beta_{11} p_1^{1/4} p_2^{1/4} p_3^{1/2} +$$

$$\beta_{12} p_1^{1/4} p_3^{3/4} + \beta_{13} p_2^{3/4} p_3^{1/4} + \beta_{14} p_2^{1/2} p_3^{1/2} + \beta_{15} p_2^{1/4} p_3^{3/4},$$

which are homogenous of degree one, constant returns to scale unit cost functions.[29]

As in Terrell (1995), the performance of the AIM(τ) is evaluated over the cubic region $\boldsymbol{\psi}^{\square} = \{\mathbf{p}: \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\}$ by defining a grid $\boldsymbol{\psi}^{\square}_{\text{g}} \subset \boldsymbol{\psi}^{\square}$ of 20 equidistant prices for each input. Thus in total $\boldsymbol{\psi}^{\square}_{\text{g}}$ consists of $Q = 20{\cdot}20{\cdot}20 = 8000$ points, $q = 1,\ldots,Q$.

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

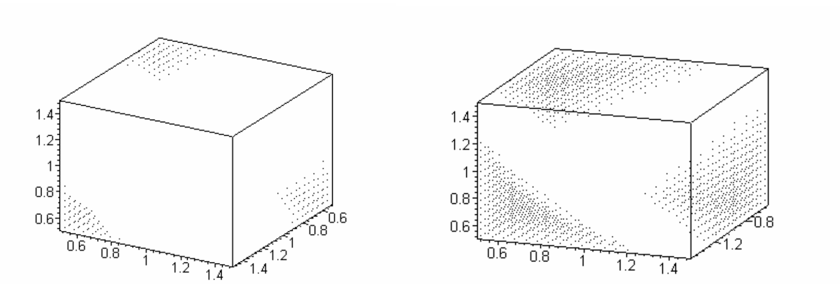
[29] A functional form is second order flexible, if it is capable of being *locally* equivalent to the true function in level, gradient, and Hessian at one given point in the price domain $\boldsymbol{\pi}$. This is the case for the AIM(1), which is equivalent to the well known Generalized Leontief.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

This grid is used to compute (a) the *maximum approximation error*, $\text{MAE}_k = \text{sgn}\{\hat{u}_{\arg\max_q\{|\hat{u}_{qk}|\},k}\}\cdot\max_q\{|\hat{u}_{qk}|\}$, and (b) the *average absolute approximation error*, $\text{AAAE}_k = Q^{-1}\sum_{q=1}^{Q}|\hat{u}_{qk}|$, 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.

2.4.1.3 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's (table 1 and 2, pp.9-10:1995) simulation experiment, and the last two columns apply the method described in section 4.

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



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

First the demand system is estimated subject to local concavity and monotonicity constraints guaranteeing regularity for the underlying AIM(τ) cost function at $\mathbf{p}^{M} = [1,1,1]$, i.e. at the mean of $\boldsymbol{\psi}^{\square}$. Compared to the other columns, the local approach provides the best model fit statistics but violates the regularity conditions in the neighbourhood of $\mathbf{p}^{M}$ (leading to regularity violations 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).

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

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

(1) Experiment based on table 1 and table 2 of Terrell (1995): True data generation process: CES technology with parameter settings a_i = 1; ρ = 0.75. In order to provide a benchmark for the average and largest error, the CES-input demand data $\mathbf{x}_k$ have, as in Terrell (1995), mean of $8000^{-1}\sum_{g=1}^{8000} x_{gk}$ = 0.2552 $\forall$ k and standard deviation of std($\mathbf{x}_k$) = 0.2230 $\forall$ k over the evaluation grid $\psi^{\square}{}_{g}$.
(2) Some considerable differences exist between our and Terrell's (1995) results. (a) Local Regularity AIM(2): Instead of 3.11% Terrell found 1.6% of concavity violations. (b) He calculated error statistics in the column 'global approach' which are about 3-4 times higher for the AAAE and 1.5 times higher for the MAE than our results: AIM(1): AAAE = 0.64146, MAE = -0.84186; AIM(2): AAAE = 0.47073, MAE = -0.63968. After careful consideration, we believe that the results in our table are the correct ones.

In the column 'global regularity' economic theory holds globally on π through the imposition of nonnegativity constraints on all the AIM parameters β (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[30] 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 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 section 3 that the model fit statistics are *always* superior to the 'mean approach', leading to a further reduction in the AAAE of 1.7% and 7.2% and to a reduction in the MAE of 8.7% and 2.3% for the AIM(1) and AIM(2), respectively.

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

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

from its appealing regularity preserving property, it seems relevant for model fit to use the mode instead of the mean.

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

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

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

2.4.2 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 $\boldsymbol{\psi}^{\square} = \{\mathbf{p}\text{: } \mathbf{p} \in \times_{k=1}^{3} [.5, 1.5]\}$, under the restriction that a) the smallest and the largest values are (again) elements of the boundary of $\boldsymbol{\psi}^{\square}$, i.e. $p_k^{\min} = 0.5\ \forall\ k$ and $p_k^{\max} = 1.5\ \forall\ k$ and that b) the points do not belong to three convex subsets that are eliminated from $\boldsymbol{\psi}^{\square}$. Suppose further that the purpose of the estimated model is to analyze $C = 4$ (policy) scenarios, and that the scenario prices are exogenously determined at 2 points within $\boldsymbol{\psi}^{\square}$ and at 2 points outside of $\boldsymbol{\psi}^{\square}$.[31] Then, a natural goal is to estimate the function such that all $N + C$ price points are regular (objective I) and that the model fit is as good as possible (objective II).

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

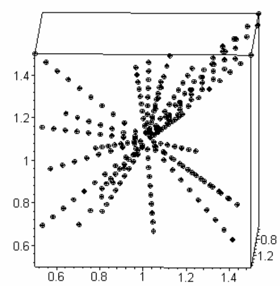
(a) $\boldsymbol{\psi}^{\square}$, as before approximated by 8000 grid points $\boldsymbol{\psi}^{\square}_{\mathrm{g}}$ and

(b) $\boldsymbol{\psi}^{\mathrm{string}} = \bigcup_{i=1}^{29} \boldsymbol{\psi}_i$, which covers all 30 = I + 1 price points by connecting 29 straight lines $\boldsymbol{\psi}_i$, $i = 1,...I$, between $\mathbf{p}^{\mathrm{M}}$ (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 $\boldsymbol{\psi}_{i\mathrm{g}}$ by taking 20 equidistant grid points between $\mathbf{p}^{\mathrm{M}}$ and the i^{th} price point, leading to a

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

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.

Fig. 4: The String grid $\boldsymbol{\psi}^{string}_{g}$



Furthermore, for the AIM(1), the grid could be further reduced to just 30 evaluation points, $\mathbf{Z}_h$, for assessing monotonicity and the sign of the first order leading principal minor. We refer to (a) as the 'cube approach' and (b) as the 'string approach'.

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

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

We also supply performance statistics for the string approach evaluated over the cube grid $\boldsymbol{\psi}^{\square}{}_{g}$. We do not necessarily advocate such an approach (i.e. defining $\boldsymbol{\psi}$ on a subset of the region where subsequent inferences will be drawn). We rather include these results[32] to again emphasize the trade off between flexibility and regularity: The regional regularity approach can become useless when $\boldsymbol{\psi}$ does not cover the empirically relevant region (because it is likely that outside of $\boldsymbol{\psi}$ 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 $\boldsymbol{\psi}$ because inferences will not be drawn from those regions.

2.5 Conclusion

In this paper we have developed 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 leads to improved model fit, and is also computationally much faster and more efficient than previous approaches while imposing both curvature and monotonicity on

[32] 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 $\boldsymbol{\psi}^{N}{}_{g}$, this is not necessarily true when evaluated over the cubic region $\boldsymbol{\psi}^{\square}{}_{g}$, (i.e. in the case of the AIM(2) the change in approximation errors are negative). The demand quantities for the out of sample prices in $\boldsymbol{\psi}^{\square} \backslash \boldsymbol{\psi}^{\text{string}}$ are calculated by (7).

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, $\boldsymbol{\psi}$, to be some connected non-convex set. We further noted that the commonly used Metropolis Hastings technique suffers from a bias of posterior density values. Finally we demonstrated that the commonly used posterior mean may be inappropriate as a point estimate. For both of the latter problems we suggested simple consistent alternatives.

It will be instructive to apply this estimation methodology empirically to estimate supply and demand systems, and other economic models requiring curvature, quasi-convexity or monotonicity restrictions. Also, it would be interesting to compare these 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 and results demonstrated in this paper promote tractability and facilitate efficiency in the analysis of regularity-preserving economic models.

Chapter 3

Can We Close the Gap Between the Empirical Model and Economic Theory? An Application to the U.S. Demand for Factors of Production

While a difficult literature, we believe that research on models permitting flexible imposition of true regularity should expand.
-- William A. Barnett & Meenakshi Pasupathy, 2003 --

3.1 Motivation

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.

[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.

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. 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] 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.

3.2 Outline of Chapter 3

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 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.

3.3 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?

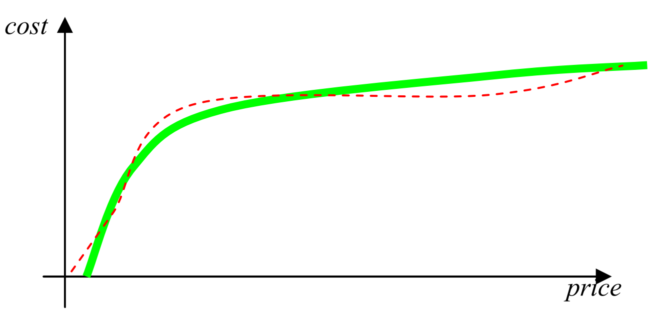
Before offering an answer, we first review some of the concepts involved. In this paper we estimate a factor demand system which is derived by Shephard's Lemma from the dual cost function $c^*(p,y,t) = \min\{p`x \mid f(x,t) \geq y\}$. Inputs are denoted by $x \in \Re^N_+$ and are transformed into output $y \in \Re^1_+$ by a smooth production function $f(x,t)$, which depends upon technological change $t \in \Re^1_+$. Duality theory implies that $c^*(z)$, where $z = [p,y,t] \in \Re_+^{2+N}$, is

HD1	: homogenous of degree one in input prices $p \in \Re^N_+$
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 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.

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

[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.

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 reasons[5]:

(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 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 evaluated in section 6.

[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).

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

3.5 Regional regularity: An alternative

3.5.1 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 $\boldsymbol{\psi} \subset \Re_+^{2+N}$ of the price × output × time space. $C_{y,p}$ and $M_{y,p}$ can be characterized by $H = 4$ vector-valued-functions $i_h(\mathbf{z};\beta)$, $h = 1,\ldots,H$, whereby the restrictions hold whenever, for a given β, $\mathbf{i}$ is nonnegative for all z in the relevant region $\boldsymbol{\psi}$,

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

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

$$i_1 = \nabla_p c \geq 0, \qquad i_2 = \nabla_y c \geq 0,$$

$$i_3 = -\operatorname{eig}[\nabla_{pp} c] \geq 0, \qquad \text{and} \qquad i_4 = \operatorname{eig}[\nabla_{yy} c] \geq 0.$$

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 $\boldsymbol{\psi}$.

> ***Definition 1:*** *The set $\boldsymbol{\psi}$ is a closed and connected subset of $\Re_+^{2+N}$ 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 $\boldsymbol{\psi}$ is defined. If $\mathbf{i} \geq \mathbf{0}\ \forall\ z \in \boldsymbol{\psi}$, we say that regularity is imposed (i) *locally* if $\boldsymbol{\psi}$ consists of one or more singular disconnected points, (ii) *globally* if $\boldsymbol{\psi} = \Re_+^{2+N}$, and (iii) *regionally* if $\boldsymbol{\psi}$ is some connected subset of $\Re_+^{2+N}$.

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 $\boldsymbol{\psi}$ is regular, it is not particularly important if the function is irregular immediately outside the boundary of $\boldsymbol{\psi}$ because inferences will not be drawn from those regions. Whereas the functions shape properties like $\mathrm{M}_{p,y}$ and $\mathrm{C}_{p,y}$ on $\boldsymbol{\psi}$ are purely determined by economic theory, the coordinates of $\boldsymbol{\psi}$ (within the right hand side variable space $\Re_+^{2+N}$) 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.

3.5.2 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

[6] Let $\boldsymbol{\Theta}$ be the K-dimensional parameter space. If all regularity conditions hold for all values of z in $\boldsymbol{\psi}$, the regular parameter set is defined as $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi} = \{\boldsymbol{\beta}\in\boldsymbol{\Theta}: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0}\ \forall\ z \in \boldsymbol{\psi}\}$, hence $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ is dependent on the choice of $\boldsymbol{\psi}$. The marginal prior on $\boldsymbol{\beta}$ is specified as an indicator function $p(\boldsymbol{\beta}|\boldsymbol{\psi}) = \mathbb{1}\{\boldsymbol{\beta} \in \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}\}$ 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, $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\psi}) \propto \int L(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})\cdot \mathbb{1}\{\boldsymbol{\beta} \in \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}\}\cdot |\boldsymbol{\Sigma}|^{-(N+1)/2}d\boldsymbol{\Sigma}$, where $L(\boldsymbol{\beta},\boldsymbol{\Sigma}|\mathbf{y})$ is the normal likelihood function.

Tan-Cruz (2000) a modified[7] Metropolis-Hastings Accept Reject Algorithm (mMHARA) is used to generate J (pseudo-) random outcomes, $\mathbf{b}^{(j)}$, $j = 1,\ldots,J$ from $p(\beta|\mathbf{y},\boldsymbol{\psi})$ on the regular support $\Theta^{\mathrm{R}}|\boldsymbol{\psi} = \{\beta\text{: } \mathbf{i}(\mathbf{p};\beta) \geq \mathbf{0}\ \forall\ \mathbf{z} \in \boldsymbol{\psi}\}$. To account for the regularity prior $\mathbb{1}\{\beta \in \Theta^{\mathrm{R}}|\boldsymbol{\psi}\}$, 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. $\mathbf{b}^{(j)} \in \Theta^{\mathrm{R}}|\boldsymbol{\psi}\ \forall\ j$. 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 $\boldsymbol{\psi}_{\mathrm{g}} \subset \boldsymbol{\psi}$ 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 $\mathbf{b}^{(j)} \in \Theta^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}\ \forall\ j$, whereby $\Theta^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}$ is the *approximated* regularity posterior support, which will tend toward the actual set $\Theta^{\mathrm{R}}|\boldsymbol{\psi}$ the finer the approximation grid $\boldsymbol{\psi}_{\mathrm{g}}$.

3.5.3 Approximating ψ

Key to the success of empirical implementation of the technique is to have 'good' construction procedures for $\boldsymbol{\psi}_{\mathrm{g}}$. The design of $\boldsymbol{\psi}_{\mathrm{g}}$ 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 $\boldsymbol{\psi}\backslash\boldsymbol{\psi}_{\mathrm{g}}$. 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 $z_i(\boldsymbol{\psi}_{\mathrm{min}})$ and $z_i(\boldsymbol{\psi}_{\mathrm{max}})$ represent the minimum and maximum of the i-th

[7] WHM identified an error in the previous MHARA literature that can bias the simulated regularity posterior. WHM suggested a simple alternative to correct for the bias.

right hand side variable. The grid is constructed by selecting $F = 10$ equidistant values for each variable: $z^f_i = z_i(\boldsymbol{\psi}_{\min}) + (f\text{-}1)F^{-1}(z_i(\boldsymbol{\psi}_{\max})\text{-}z_i(\boldsymbol{\psi}_{\min}))\ \forall\ f \in \{1,2,\ldots, F\}$ and using all possible $Q = 10^6 = F^{\dim(z)}$ combinations to generate the Q-grid $\boldsymbol{\psi}^{\square}{}_{\mathrm{g}} \subset \boldsymbol{\psi}^{\square}$. 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 $\forall\ \mathbf{z} \in \boldsymbol{\psi}^{\square}$. 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 $316 = 1+(F\text{-}1)(T+S)$ 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.

3.5.4 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 $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ is not a convex set. Instead using the mode $\boldsymbol{\beta}^{(\text{mode})}$ = $\underset{\boldsymbol{\beta}\in\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_{\mathrm{g}}}{\arg\max}\left\{p(\boldsymbol{\beta}\,|\,\mathbf{y},\boldsymbol{\psi}_{\mathrm{g}})\right\}$ guarantees that the point estimate resides in the regular support $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$. In order to approximate the solution based upon the MCMC outcomes $\{\mathbf{b}^{(j)}\}_{j=1}^{J}$, one can simply compare the values $p_{\mathrm{u}}(\mathbf{b}^{(j)}|\mathbf{y})$ $\forall$ j resulting from the mMHARA as $\mathbf{b}^{(\text{mode})} = \underset{\mathbf{b}^{(j)}}{\operatorname{argmax}}\left\{|\,(N-L)\boldsymbol{\Sigma}(\mathbf{b}^{(j)})\,|^{-N/2}\right\}$. 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 $\boldsymbol{\beta}^{(\text{mode})}$ if, as we have done above, an uninformative prior distribution on $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}$ is employed (see footnote 5). Here we prefer the Bayesian interpretation because finite sample confidence intervals and standard errors of functions of $\boldsymbol{\beta}$ 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 $z \in \boldsymbol{\psi}$. Also since

$\beta^{(mode)}$ could lie on the boundary of $\Theta^R|\psi$ further complications arise (see Geweke 1986, Andrews 1999, 2001).

3.6 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 CO_2 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).[8]

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.[9]

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 $\boldsymbol{\psi}$ sets.

---

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

9 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	p_K	p_L	p_E	p_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).

3.6.1 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

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

$$+ y^2\,\Sigma_{i=1}^{N} \beta_i p_i + y t^2\,\Sigma_{i=1}^{N} \gamma_i p_i$$

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.[10] Hence the N estimated equations are

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

[10] 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 mMHARA simulator.

It is assumed that the $T \times 1$ error vectors $\mathbf{u}_n$, $n = 1,\ldots,4$ are contemporaneously correlated, such that the estimating equations can be written in form of the Gaussian seemingly unrelated regression (SUR) system. Finally $t = 1,2,\ldots T$. 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 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		$\psi = \varnothing$:	$M_{p,y}$ 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		$\psi = \Re_+^{2+N}$	Nonnegative orthant of all right hand side variables of *z*.
Regional	Cube[11] approach	$\Psi^{\square p}_1 = \{\mathbf{p}\text{: } \mathbf{p} \in \times_{k=1}^{3} [1.0, 1.5]\}$	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.*
Regional	Cube[11] approach	$\Psi^{\square p}_2 = \{\mathbf{p}\text{: } \mathbf{p} \in \times_{k=1}^{3} [0.5, 3.0]\}$	This set was chosen by Terrell (1996). It ensures that $\psi^{\square p}_2$ covers all observed prices and beyond.*
Regional	Cube[11] approach	$\Psi^{\square z}_i = \{\mathbf{z} \in \psi^{\square p}_i \times [y_{\min}, y_{\max}] \times [t_{\min}, t_{\max}]\}$	$\psi^{\square z}_i$, $i \in \{1,2\}$: These sets expand $\psi^{\square p}_i$ over the remaining dimensions *t* and *y*.
Regional	String approach	$\psi^{\text{string}} = \bigcup_{t=1}^{26} \psi_i$	ψ^{string} covers 26 = T+1 points in $\Re_+^{2+N}$ by connecting straight lines ψ_i between the right hand side variables mean, $\mathbf{z}^{\mathrm{M}}$, and each of the T observations. We approximate each line ψ_i by ψ_{ig} by taking 10 equidistant grid points between $\mathbf{p}^{\mathrm{M}}$ 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*.

[11] In this paper all grid sets $\psi^{\square ''}_{ig}$ 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 $V(\Psi^{\square p}_1) = 0.0625$ to $V(\Psi^{\square p}_2) = 150.0625$. As an alternative one could express the number of grid points as a function of the volume, $F(V)$.

Columnwise table 3 displays the various approaches imposing restrictions on differently designed sets $\boldsymbol{\psi}$, which definitions are first provided in table 2. In Table 3, the symbols in parentheses indicate which shape conditions are imposed. For example, $\boldsymbol{\psi}^{\square p}{}_{2}(\mathrm{M}_{p},\mathrm{C}_{p})$ 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 $\boldsymbol{\psi}^{\square}{}_{1}$ and $\boldsymbol{\psi}^{\square}{}_{2}$ 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.

Table 3: Generalized Leontief Input Demand System, estimated by 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)		Ø('none')	$\Re^{2+N}(C_y)$		$\Re^{2+N}$('all')	z_1('all')		$\psi^{\varphi}{}_1(C_y,M_y)$		$\psi^{\varphi}{}_2(C_y,M_y)$		$\psi^{\infty}{}_1$('all')		$\psi^{\infty}{}_2$('all')		ψ^{string}('all')	
evaluated at	% of shape violations at grid points	mode	mean	mode	mode (b^{σ})	mean	mode	mean	mode	mean	mode	mean	mode	mean	mode	mean	mode
$\psi^{o}{}_1$	C_y 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_y 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_y 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_y 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
$\psi^{o}{}_2$	C_y 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_y 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_y 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_y 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 *Generalzed 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 $p(\mathbf{b}|y, Ø)$ 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.

3.6.1.1 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 $\boldsymbol{\psi}^{\square}$ leading to, among other things, a failure of the law of demand.[12] 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).

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.

[12] For comparison, this estimation exactly repeats the unrestricted estimation of Diewert and Wales (1987: table II) and TERRELL's (table 3: 1996).

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 $\Theta^{R}|\Re_{+}^{2+N} \subset \Theta^{R}|\boldsymbol{\psi}$. 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.

3.6.1.3 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 $\boldsymbol{\psi}^{\square}_{2}$. 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 $\boldsymbol{\psi}^{\square}_{1}$ does not cover all the relevant data points (compare the min/max values in table 1 and the definition of $\boldsymbol{\psi}^{\square}_{1}$ 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

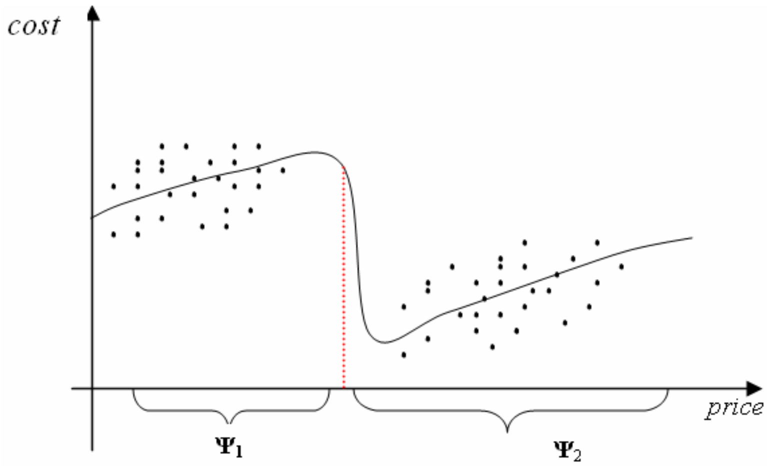


Illustration of a cost function partially regular within ψ_1 and within ψ_2. The overall model is irregular on the domain $\psi = \psi_1 \cup \psi_2$ although only one grid point is violated (here indicated on the domain by the red ray). Overall regularity is violated because cost must not decrease with 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.

3.6.1.4 Regional Regularity

We first replicate Terrell's (1996) estimation, i.e. using the Gibbs accept reject simulator and the same $\boldsymbol{\psi}$-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 $\boldsymbol{\psi}^{\square z}{}_1$. In contrast, on the domain $\boldsymbol{\psi}^{\square z}{}_2$, 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 mMHARA simulator. This method effectively imposes all regularity conditions on any set $\boldsymbol{\psi}$ of interest. Hence, as shown in the columns 6 and 7 regularity holds in $\boldsymbol{\psi}^{\square z}{}_1$ and $\boldsymbol{\psi}^{\square z}{}_2$. 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 $\mathbf{b}^{mean} \notin \Theta^R|\boldsymbol{\psi}^{\square}{}_2$ is due to the fact that the regular parameter support $\Theta^R|\boldsymbol{\psi}^{\square z}{}_2$ is not a convex set and demonstrates the importance of the proposition outlined in section 5.4. Instead our preferred point estimate, the mode $\mathbf{b}^{\text{mode}} \in \Theta^{R}|\boldsymbol{\psi}^{\square}{}_{2}$ and it guarantees regional regularity within the domain $\boldsymbol{\psi}^{\square}{}_{2}$. 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 $\boldsymbol{\psi}^{\square z}{}_{2}$, 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 $\boldsymbol{\psi}^{\square}$. A motivation for the KLEM data set to investigate in non convex sets for $\boldsymbol{\psi}$ is probably best described by Gallant and Golub, 1984: *'The exogenous variable* [p_t and y_t] *for t* = 1947,…,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 (~) with the two end-points and the middle elevated.*' This description indicates that constructing $\boldsymbol{\psi}^{\square}$ 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'.[13] Comparing the posterior mean of the string approach versus the mean of column 6 implies that here mMHARA samples from a regular parameter superset $\Theta^{R}|\boldsymbol{\psi}^{\text{string}} \supset \Theta^{R}|\boldsymbol{\psi}^{\square z}{}_{1}$, 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%.

[13] It is straightforward to incorporate other out of sample areas into the construction of $\boldsymbol{\psi}$, i.e. in order to obtain more general domains relevant for forecasting purposes. See WHM for details.

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 extend ‘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. [14]

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

[14] 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.

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 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.

3.6.2 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 mMHARA 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.

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

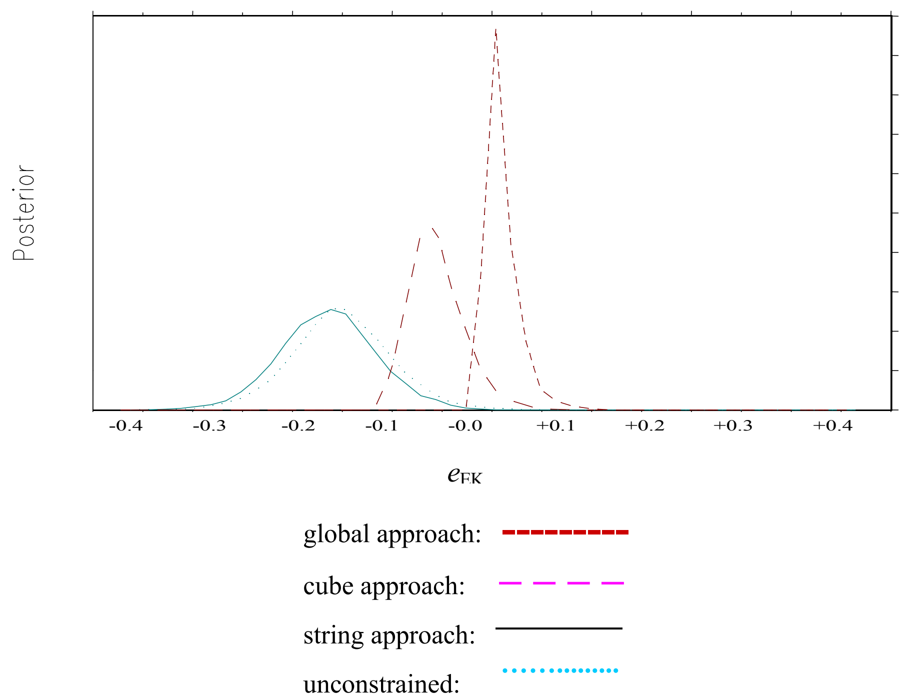


Table 4: Price elasticities matrices at 1947

Unrestricted SUR

	K	L	E	M
mode estimates				
K	-0.0974	0.4461	-0.1315	-0.2172
L	0.0921	-0.1774	0.0922	-0.0069
E	-0.1579	0.5362	-0.6167	0.2384
M	-0.0168	-0.0026	0.0154	0.0040
mean estimates				
K	-0.0978	0.4384	-0.1320	-0.2086
L	0.0905	-0.1909	0.0926	0.0078
E	-0.1585	0.5384	-0.6136	0.2337
M	-0.0162	0.0029	0.0151	-0.0018
std. of posterior distributions				
K	0.0625	0.1576	0.0422	0.1998
L	0.0325	0.2355	0.0402	0.2393
E	0.0507	0.2338	0.1284	0.2067
M	0.0155	0.0898	0.0133	0.0994

Local Approach

	K	L	E	M
mode estimates				
K	-0.1488	0.2897	-0.1412	0.0004
L	0.0598	-0.4092	0.0888	0.2606
E	-0.1696	0.5163	-0.6659	0.3192
M	0.0000	0.0978	0.0206	-0.1184
mean estimates				
K	-0.1485	0.2896	-0.1412	0.0002
L	0.0598	-0.4110	0.0888	0.2624
E	-0.1696	0.5160	-0.6655	0.3191
M	0.0000	0.0985	0.0206	-0.1191
std. of posterior distributions				
K	0.0492	0.1228	0.0415	0.1422
L	0.0253	0.1751	0.0392	0.1689
E	0.0498	0.2280	0.1203	0.2006
M	0.0110	0.0634	0.0129	0.0665

Global concavity

	K	L	E	M
mode estimates				
K	-0.1369	0.0685	0.0145	0.0539
L	0.0141	-0.3349	0.0569	0.2638
E	0.0174	0.3307	-0.5425	0.1943
M	0.0042	0.0990	0.0125	-0.1157
mean estimates				
K	-0.1364	0.0682	0.0145	0.0537
L	0.0141	-0.3334	0.0570	0.2624
E	0.0174	0.3311	-0.5431	0.1946
M	0.0042	0.0984	0.0126	-0.1151
std. of posterior distributions				
K	0.0502	0.0437	0.0141	0.0417
L	0.0090	0.1712	0.0289	0.1676
E	0.0169	0.1681	0.1307	0.1208
M	0.0032	0.0629	0.0078	0.0638

Regional approach on cube1

	K	L	E	M
mode estimates				
K	-0.1488	0.2897	-0.1412	0.0004
L	0.0598	-0.4092	0.0888	0.2606
E	-0.1696	0.5163	-0.6659	0.3192
M	0.0000	0.0978	0.0206	-0.1184
mean estimates				
K	-0.1672	0.2300	-0.1312	0.0683
L	0.0475	-0.3898	0.0831	0.2592
E	-0.1575	0.4830	-0.6566	0.3310
M	0.0053	0.0973	0.0214	-0.1239
std. of posterior distributions				
K	0.0468	0.1046	0.0389	0.1239
L	0.0216	0.1721	0.0339	0.1720
E	0.0467	0.1971	0.1109	0.1864
M	0.0096	0.0645	0.0120	0.0691

Regional approach on cube2

	K	L	E	M
mode estimates				
K	-0.1605	0.1938	-0.0585	0.0251
L	0.0400	-0.3584	0.0915	0.2268
E	-0.0703	0.5321	-0.6585	0.1966
M	0.0019	0.0851	0.0127	-0.0997
mean estimates				
K	-0.1613	0.1237	-0.0398	0.0774
L	0.0255	-0.3661	0.0691	0.2715
E	-0.0477	0.4015	-0.5641	0.2103
M	0.0060	0.1019	0.0136	-0.1214
std. of posterior distributions				
K	0.0483	0.0729	0.0228	0.0781
L	0.0151	0.1599	0.0330	0.1575
E	0.0274	0.1921	0.1100	0.1610
M	0.0061	0.0591	0.0104	0.0616

String approach

	K	L	E	M
mode estimates				
K	-0.1488	0.2897	-0.1412	0.0004
L	0.0598	-0.4092	0.0888	0.2606
E	-0.1696	0.5163	-0.6659	0.3192
M	0.0000	0.0978	0.0206	-0.1184
mean estimates				
K	-0.1474	0.2900	-0.1429	0.0004
L	0.0599	-0.4215	0.0886	0.2730
E	-0.1716	0.5153	-0.6714	0.3277
M	0.0000	0.1024	0.0211	-0.1236
std. of posterior distributions				
K	0.0475	0.1246	0.0406	0.1443
L	0.0257	0.1748	0.0397	0.1725
E	0.0487	0.2308	0.1204	0.1992
M	0.0112	0.0647	0.0129	0.0698

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

		Global concavity approach			String approach		
	year	(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)

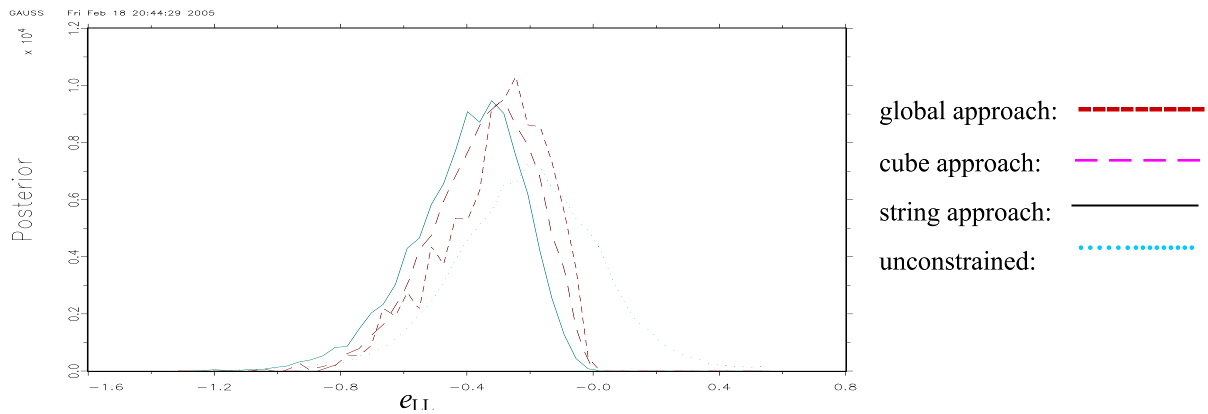
In the first row calculations are based on the year 1947. This represents our results comparably to the tables and figures provided in previous studies (such as Berndt and Wood (1979), Diewert and Wales (1984), Barnett et al. (1991) and Terrell (1996)). Numbers in parenthesis provide the 90% coverage probability intervals of respective changes and elasticities. Intervals are computed by using 100,000 mMHARA 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 mMHARA 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.[15] 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 $\boldsymbol{\psi}$, 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.

[15] 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.

3.6.3. Does the KLEM data set support duality theory?

3.6.3.1 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 $\mathbf{b}^{rr}$ of the string approach (since $\mathbf{b}^{rr}$, (column 8) satisfies all shape restrictions, the model $c(z,\mathbf{b}^{rr})$ 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).[16] 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.

[16] 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.

3.6.3.2 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 $\mathbf{\Theta}^{R}|\Re_{+}^{2+N} \subset \mathbf{\Theta}^{R}|\boldsymbol{\psi}$. 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 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 mMHARA estimator closes this gap.

3.7 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.

Chapter 4

Daylight Time and Energy
Evidence from an Australian Experiment

I say it is impossible that so sensible a people…should have lived so long by the smoky, unwholesome, and enormously expensive light of candles, if they had really known, that they might have had as much pure light of the sun for nothing. – Benjamin Franklin, 1784 –

4.1 Introduction

One principal socio-economic problem is the optimal allocation of individuals' activities—sleep, work, and leisure—over the twenty-four hours of the day. In today's world of artificial lighting and heating, people set their active hours by the clock rather than by the natural cycle of dawn and dusk. In one of the earliest statistical treatments in resource economics, *An Economical Project*, Benjamin Franklin (1784) criticizes this behavior because it wastes valuable sources of morning daylight and requires expensive candles to illuminate the nights. Franklin calculates that this misallocation causes Paris to consume an additional 64 million pounds of tallow and wax annually.

Governments have also recognized this resource allocation problem, and have attempted to address it through the mechanism of Daylight Saving Time (DST).[1] Each year we move our clocks forward by one hour in the spring, and adjust them back to Standard Time in the fall. Thus, during the summer, the sun appears to set one hour later and the "extra" hour of evening daylight is presumed to cut electricity demand.

---

[1] Historically, DST has been most actively implemented in times of energy scarcity. The first application of DST was in Germany during World War I. The U.S. observed year-round DST during World War II and implemented several substantial extensions during the energy crisis in the 1970s (*Emergency Daylight Savings Time Energy Conservation Act*, 1973). Today, DST is observed in over seventy countries worldwide. But DST is also heavily criticized for the inconveniences it creates on the days when the switch between DST and Standard Time occurs. For more information on the history of DST, see the recent books by Prerau (2005) and Downing (2005); Beauregard-Tellier (2005) provides an overview on the recent DST-energy literature.

Today, heightened concerns regarding energy prices and the externalities of fossil fuels are driving renewed interest in implementing DST in many countries.[2] The United States recently passed legislation to extend DST by one month with the specific goal of reducing electricity consumption by 1% during the extension (*Energy Policy Act*, 2005). Beginning in 2007, the U.S. will switch to DST in March rather than in April. California is considering even more drastic changes—year-round DST and double DST—that are predicted to save up to one billion U.S. dollars annually (*Joint Senate Resolution*, 2001).

Our study challenges the DST-energy literature findings that have been directly used to justify these calls for the expansion of DST. Across the studies we surveyed, estimates of an extension's effect on electricity demand range from 0.6% to 3.5%. The most widely cited savings estimate of 1% is based on an examination of a U.S. extension to DST that occurred in response to the Arab oil embargo (DOT, 1975). Due to the age of this study, it is possible that its findings are no longer applicable today—for example, because the widespread adoption of air conditioning has altered intraday patterns of electricity consumption.

One primary method for predicting the effects of DST on electricity use is to employ simulation models, such as the 2001 study by the California Energy Commission (CEC) that is being used to promote year-round DST in California.[3] It

---

[2] Non-U.S. regions currently considering extending DST are Japan, Canada, New Zealand, and Australia with cited electricity savings of 2.2%-3.5%. The purpose for the extension plans differ by country and range from cutting greenhouse gas emissions by 440,000 tons of CO_2 in Japan to conserving water resources used in New Zealand to generate hydropower. For details see ECCJ (2006), Young (2005), Eckhoff (2001) and Hansard (2005).

[3] Until today, the DST system proposed in California's *Joint Senate Resolution* (2001) has not been implemented. *"Congress and the White House did not act on the request because of the world-changing events of September 11,*

predicts three benefits: (1) a 0.6% reduction in electricity consumption, (2) lower electricity prices, driven by a reduction in peak demand, and (3) a lower likelihood of rolling blackouts. However, this study is not based on firm empirical evidence, instead it uses electricity market data under the current DST scheme to simulate demand under extended DST. It may therefore fail to capture the full behavioral response to a change in DST timing.[4]

An alternative approach is to examine electricity consumption a week before and a week after currently existing springtime changes. The set of studies taking this approach forecast larger drops in electricity use: from 2.2% in Ontario, Canada (Young, 2005) up to 3.5% in New Zealand (Eckhoff, 2001). However, the week after the springtime change has longer and warmer days which, even in the absence of DST, would change electricity consumption, potentially biasing the studies' results.

In our study, we offer a new test of whether extending DST decreases energy consumption by evaluating a quasi-experiment that occurred in Australia in 2000. Typically, three of Australia's six states observe DST beginning in October (which is seasonally-equivalent to April in the northern hemisphere). However, to facilitate the 2000 Olympics in Sydney, (located in New South Wales), two of these three states began DST two months earlier than usual. Because the Olympics can directly affect the electricity demand we focus on Victoria—which did not host Olympic events—as the treated state, and use its neighbor state, South Australia, which did not extend

---

2001" (Aldrich, 2006). Subsequently, the federal *Energy Policy Act* has been considered more urgent, rather than changing DST state by state.

[4] We found one more study by Rock (1997). Using a complex simulation model he finds that year-round DST decreases demand by 0.3% and electricity expenditures decrease by 0.2%. However, the simulation does not include non-residential electricity use, which accounts for 74% of U.S. total electricity consumption (EIA, 2005).

DST, as a control. Furthermore by dropping the two week long Olympic period from the two month treatment period we remove confounding effects. Using a detailed panel of half-hourly electricity consumption and prices, as well as the most detailed weather information available, we examine how the DST extension affected electricity demand in Victoria. This experimental setting and rich dataset obviate the need to rely on simulations in our study.

Our treatment effect estimation strategy is based on a difference in differences (DID) method that exploits, in both the treatment state and the control state, the difference in demand between the treatment year and the control years. We augment the standard DID model in several innovative ways. Most notably, we take advantage of the fact that DST does not affect electricity demand in the afternoon; we can therefore use changes in relative afternoon consumption to control for unobserved shocks that are not related to DST. We show that this allows us to employ a much more relaxed identifying assumption compared to the standard DID setting.

Our results show that the extension failed to conserve electricity. The point estimates suggest that energy consumption increased rather than decreased, and that the within-day usage pattern changed substantially, leading to a high morning peak load. The morning wholesale electricity prices therefore increased sharply. These results contradict the DST-benefits claimed in the prior literature.

We further analyze whether the prior approaches to forecasting electricity demand could have predicted the outcomes of the Australian experiment. This is a relevant question for many countries that wish to evaluate the benefits of an extension.

We find that both the simulation model used in California and the "week before / week after" technique produce estimates that are biased in the direction of energy savings, which casts suspicion on the models' previous policy recommendations.

Finally, it should be noted that Australia—ranked highest in greenhouse gas (GHG) emissions per capita worldwide—is currently debating whether to permanently extend DST in a manner similar to that done in 2000 (Turton and Hamilton, 2001; Hansard, 2005).[5] Our results indicate that the claims that extending DST in Australia will significantly decrease energy use and GHG emissions are at best overstated, and at worst carry the wrong sign. Also, while we cannot apply our results to other countries without adjustment for behavioral and climatic differences, this study raises concern that the U.S. is unlikely to see the expected energy conservation benefits from extending DST.

The remainder of this paper is organized as follows: the next section provides a brief overview of the DST system in Australia and the changes that occurred in the year 2000. After describing our dataset and presenting preliminary graphical results, section 4 discusses our exogeneity assumption and the treatment effect estimation strategy. Section 5 presents the empirical findings. In section 6, we provide an overview of the two methods previously used to analyze the effects of extending DST on energy use. Section 7 and 8 then discuss the application of these two methods to Australia. We conclude by summarizing our main results and provide policy implications.

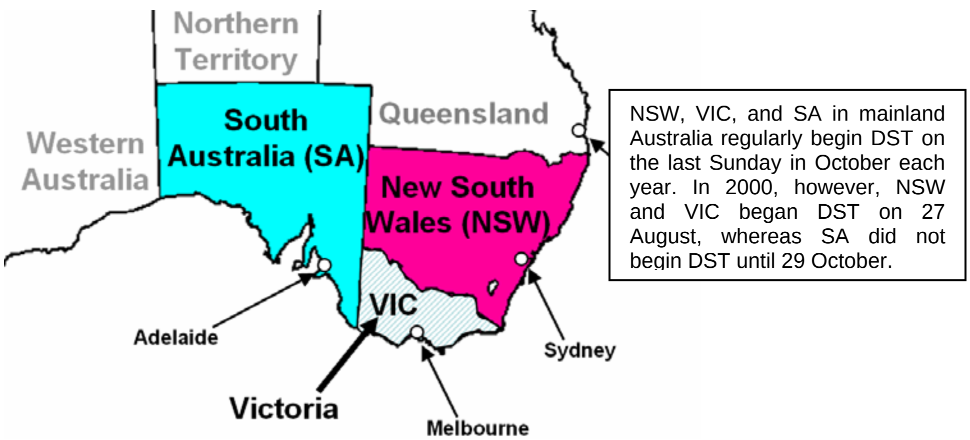
---

[5] In Australia, 92% of the electricity produced relies on the burning of fossil fuels, which substantially contributes to the GHG emissions.

4.2 Background on Daylight Saving Time in Australia

The geographical area of interest is the Australian continent's eastern part, displayed in Figure 1. Three states in the south east of the mainland observe DST—South Australia (SA), New South Wales (NSW) and Victoria (VIC). DST typically starts on the last Sunday in October and ends on the last Sunday in March. Queensland, the Northern Territory and Western Australia do not observe DST. Table 1 provides summary statistics and geographical information for the capitals of these states, where the populations and electricity demand are concentrated.

Figure 1: East Australia, states and major cities



In 2000, NSW and VIC started DST two months earlier than usual—on 27 August instead of 29 October—while SA maintained the usual DST schedule. The extension was designed to facilitate the seventeen days of the Olympic Games that took place in Sydney, in the state of NSW, from 15 September to 1 October.

Three rationales for the extension were put forward in 1991 when Sydney applied to the International Olympic Committee (Hansard, 1999a and 1999b).

(a) Many afternoon Olympic events ended near 17:30, and evening events began between 18:00 and 19:00. Extended DST would allow the movements of visitors to and from stadia to take place in sunlight rather than twilight. This was expected to improve the visitors' well-being by providing higher temperatures, more daylight, and better security.

Table 1: Geographic and population characteristics in east Australia

State Capital	State	State Income/Capita in 1000 AUD	City Population in millions	State Population in millions	Latitude South	Longitude East	Sunrise	Sunset
Sydney	NSW	41.4	4.3	6.5	33°5'	151°1'	5:50	17:45
Melbourne	VIC	39.3	3.7	4.8	37°47'	145°58'	6:20	18:10
Adelaide	SA	33.4	1.1	1.5	34° 55'	138° 36'	6:50	18:35

All estimates are of 2000. Sunrise and sunset hours refer to Eastern Australian Standard Time in the month of September. Additional astronomical data are detailed in Appendix A.

(b) Extended DST would reduce on-field shadows on the playing fields that could hinder both athletes and television broadcasting quality.

(c) Due to wind and weather conditions in September, rowing would need to start at 7:30am under Standard Time. DST permits rowing to start at 8:30am, to the benefit of spectators.

Notably, none of the justifications for the DST extension were related to energy usage.

A timeline of events is displayed in Figure 2. The decision to start DST three weeks prior to the beginning of the Olympic Games was intended to avoid confusion for athletes, officials, media and broadcasters and other international visitors who would likely arrive prior to the opening of the games. The opening of the Olympic village was scheduled for 3 September 2000. VIC adopted the NSW timing proposal to avoid inconveniences for those living on the NSW-VIC border (see Figure C1 in

Appendix C). However, SA did not extend DST in 2000 due to the opposition of the rural population (Hansard 1999a, 1999b, 2005).

Figure 2: Timeline of 2000 events in New South Wales, Victoria and South Australia

Table 2: Summary statistics of data used from 1999 to 2001, 27 August to 27 October

State	Summary of all years, 8928 observations per state						Summary by year, Olympic dates excluded					
							1999		2000		2001	
	Variable	[unit]	mean	std	min	max	mean	std	mean	std	mean	std
Victoria	Demand	[MW]	5253.68	550.56	3777.31	6861.32	5153.71	526.74	5331.40	562.57	5403.20	570.17
	Price	[AUD/MWh]	27.36	97.20	-305.78	4527.21	19.72	6.37	45.09	187.72	29.55	88.02
	Temperature	[Celsius]	12.88	4.26	2.15	27.30	13.61	4.56	11.75	3.71	12.24	3.84
	Precipitation	[mm/hour]	0.08	0.48	0.00	15.40	0.07	0.52	0.15	0.76	0.04	0.24
	Wind	[meter/sec]	5.11	3.09	0.00	18.75	4.84	2.99	5.47	2.82	4.88	2.70
	Pressure	[hPa]	1015.23	7.61	990.30	1031.95	1017.81	6.40	1011.44	7.21	1011.93	6.33
	Sunshine	[hours/day]	6.29	3.65	0.00	12.20	6.76	3.85	5.81	3.61	5.72	3.57
	Humidity	[RH%]	71.00	17.18	19.00	101.50	70.38	16.73	73.36	15.65	71.70	17.55
	Employment	[in 1000]	2254.21	43.67	2154.81	2303.30	2192.68	14.71	2271.98	12.53	2289.37	11.92
	Non-Working Day	[% of days]	0.44	0.50	0.00	1.00	0.29	0.46	0.29	0.46	0.27	0.44
	School-Vacation	[% of days]	0.16	0.37	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Holiday	[% of days]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
South Australia	Demand	[MW]	1368.48	196.92	909.87	1954.26	1339.32	179.99	1397.31	202.81	1424.02	203.35
	Price	[AUD/MWh]	41.06	120.75	3.50	5000	55.66	167.16	52.57	168.27	27.69	18.63
	Temperature	[Celsius]	14.91	4.24	4.05	31.60	15.76	4.87	14.08	3.20	13.66	3.25
	Precipitation	[mm/hour]	0.07	0.38	0.00	7.60	0.00	0.00	0.13	0.56	0.12	0.48
	Wind	[meter/sec]	4.54	2.76	0.00	17.00	4.28	2.58	5.23	2.87	4.69	2.78
	Pressure	[hPa]	1016.21	6.91	989.95	1030.80	1017.81	6.73	1014.18	7.01	1013.41	6.45
	Sunshine	[hours/day]	7.39	3.44	0.00	12.40	8.52	3.10	7.22	3.43	6.48	3.41
	Humidity	[RH%]	66.40	18.41	9.00	98.00	62.73	19.24	69.06	16.45	70.00	17.38
	Employment	[in 1000]	679.28	7.33	662.94	687.75	668.83	2.80	684.35	2.50	682.81	2.42
	Non-Working Day	[% of days]	0.45	0.50	0.00	1.00	0.34	0.47	0.29	0.46	0.39	0.49
	School-Vacation	[% of days]	0.16	0.37	0.00	1.00	0.05	0.22	0.00	0.00	0.12	0.33
	Holiday	[% of days]	0.02	0.12	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00

Abbreviations: MW = Megawatts; AUD/MWh = Australian Dollars per Megawatt-hour; mm = millimeters; hPa = Hectopascal; RH% = relative humidity %. Note that the maximum wholesale electricity price is capped at 5000 AUD/MWh from 1999-2000, and at 10,000 AUD/MWh in 2001. The cap is designed to mitigate generator market power (NEMMCO, 2005).

4.3 The Australian data and graphical results

4.3.1 Data

Electricity consumption and price data are obtained from Australia's National Electricity Market Management Company Limited (NEMMCO). These consist of half-hourly electricity demand and prices by state from 13 December, 1998 to 31 December, 2005.

Because electricity demand is heavily influenced by local weather conditions, we use two datasets from the Bureau of Meteorology at the Australian National Climate Centre. The first consists of hourly weather station observations in Sydney, Melbourne, and Adelaide—the 3 cities that primarily drive electricity demand in each state. The data cover 1 January, 1999 to 31 December, 2005 and include temperature, wind speed, air pressure, humidity and precipitation. The second dataset consists of daily weather observations, including the total number of sunshine hours per day.

We also collected information regarding state-specific holidays and public school vacations to control for their effect on electricity usage. We identify "transition vacation days" as working days sandwiched between a holiday and a weekend. For example, the Melbourne Cup in Victoria is on the first Tuesday of November each year. Because many employees take an extended weekend vacation, we model the Monday as a transition vacation day.

Table 2 provides summary statistics for each of these variables by state in the period during the DST extension, end of August to end of October, as well as for the treatment period in 2000 and the adjacent years 1999 and 2001. Displayed are the mean, standard deviations (std), minimum (min) and maximum (max). More details on the entire dataset as well as on our procedures for dealing with missing data are provided in Appendix B.

4.3.2 The impact of the DST extension on electricity consumption and prices

The goal of the empirical analysis is to examine the effect of the extension of DST on electricity use and prices. Before presenting the econometric model, the main intuition can be obtained by the graphical analysis presented in Figure 3. Panel (a) displays the average half-hourly electricity demand in Megawatts (MW) in the control state of SA during the treatment period[6], in 1999, 2000, and 2001, and the panel (b) shows the same for VIC. SA's load shape is very stable over these three years, featuring an increase in consumption between 05:00 and 10:00, a peak load between 18:00 and 21:00, and then a decrease until about 04:00 on the following morning. In particular, SA's demand in 2000 appears unaffected by the DST extension in its neighbors VIC and NSW.[7]

In VIC, however, the 2000 load shape is quite different from the loads in 1999 and 2001—the treatment dampens evening consumption, but leads to higher morning peak demand. This behavior is consistent with the expected effects of DST's one-hour time shift: less lighting and heating are required in the evening, and more in the

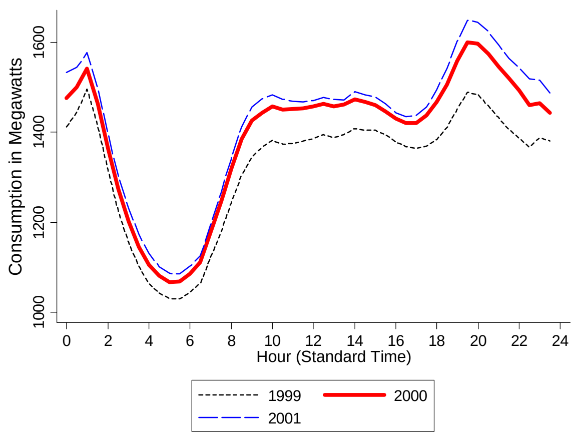
---

[6] The treatment period covers 27 August, 4am to 27 October, excluding 15 September to 2 October. This corresponds to the extension period in 2000, less the 17 days of the Olympic games.

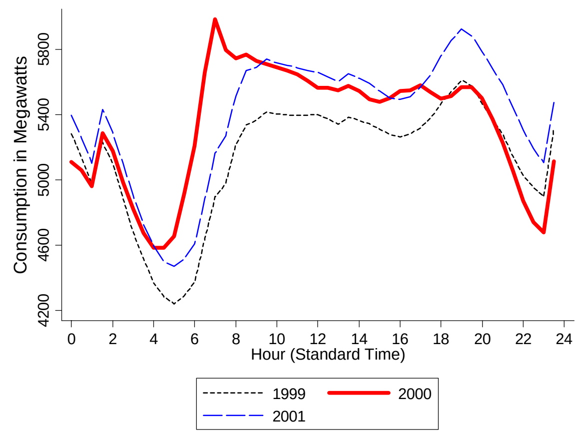
[7] Hamermesh et al. (2006) examine spatial coordination externalities triggered by time cues. Their results imply that SA in 2000 may have adjusted its behavior in response to the treatment in VIC. In particular, their model predicts that people in SA would awaken earlier in the morning to benefit from aligning their activities with their neighbors in VIC. However, the effects that Hamermesh et al. calculate are small, and Panel (a) of Figure 3 does not show evidence of such a time shift.

morning—particularly from 07:00 to 08:00—driven by reduced sunlight and lower temperatures. During the treatment period, the latest sunrise in Melbourne (on 27 August) occurs at 07:51, and the average sunrise occurs at 06:55. Further, the 07:00 to 08:00 interval is the coldest hour of the day; the average temperature for this hour is only 9°C. The one-hour clock time shift imposed by DST causes people to awaken in cold, low light conditions. This causes an increase in electricity demand that persists even one hour after sunrise.

Figure 3: "September and October" average half hourly electricity demand in South Australia (control) and Victoria (treated in 2000)[8]



(a) South Australia (control)



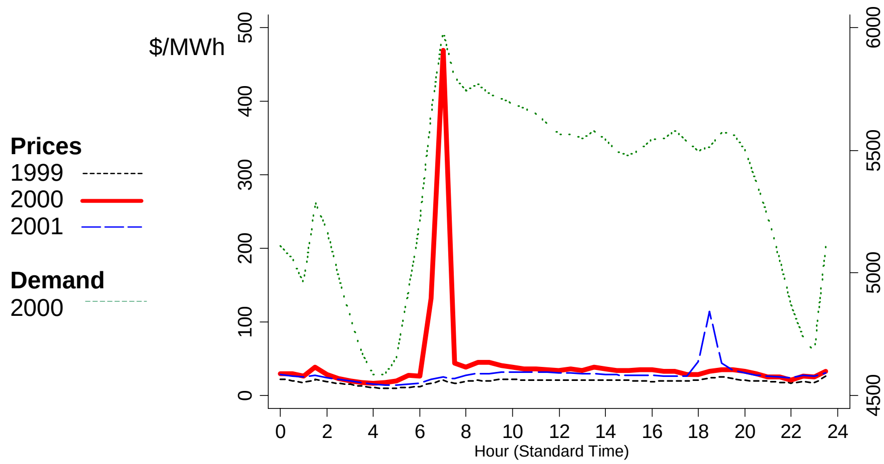
(b) Victoria (treated in 2000)

Panel (b) also casts doubt on the claims that extended DST brings additional benefits in the form of higher system reliability due to a more balanced load shape (for a discussion on these benefits see CEC, 2001). While the extension does reduce the evening peak load in VIC in 2000, it creates a new, sharp peak in the morning. This

[8] The "zig-zag" pattern that occurs between 11pm and 2am in both states is due to centralized off-peak water heating that is activated by automatic timers (Outhred, 2006). The yearly increases in electricity demand can be attributed to population growth (2% in VIC and 1% in SA) and state specific economic conditions—the real gross state income per capita grew by 3% and 5% in VIC and SA respectively. Despite these level shifts, the load patterns are remarkably similar for the control years.

2000 morning peak is even higher than the evening peak in 2001, and its sharp increase and decrease around 07:00-8:00 are steeper than those for any peak period found elsewhere in our data set.

Figure 4: "September and October" average half hourly electricity prices and demand in Victoria (treated state in 2000)



Our preliminary analysis also casts doubt on the claims that extending DST brings additional benefits in the form of reduced electricity prices. [9] As shown in Figure 4, the new morning peak in demand is coincident with a large spike in wholesale prices. Morning price spikes occurred on every working day during the first two weeks of the extension, suggesting that the generation system was initially stressed to cope with the steep ramp in demand.

[9] The fixed costs of electricity generation are largely determined by the peak load. Econometric studies suggest that higher peak loads, relative to the average load, increase average costs significantly (e.g. Filippini and Wild, 2000). The intuition for this is that when the load shape is flat, supply can be generated by coal-fired base-load generators with low variable costs. Volatile load shapes, however, require natural gas and oil-fired generators which can quickly ramp up or down, but have higher variable costs. Characteristics of the different generators used in Australia, their warm up times, supply costs, environmental impacts and the market mechanism to determine the wholesale prices of Figure 4 are further detailed in Appendix C.

The answer to the central question of whether the evening decrease outweighs the morning increase, or vice-versa, is, however, not clear from this cursory analysis since it does not account for important determinants, such as economic conditions, school vacations, weather and other factors that change over time. To obtain the unconfounded effect of the treatment, we employ regression analysis, as described in the following sections. The variables used to undertake this are displayed in Table 2.

4.4 Empirical Strategy for measuring the effect of DST on electricity use

The following two subsections describe our empirical strategy to identify (4.1) and quantitatively measure (4.2) the effect of extending DST on electricity use.

4.4.1 Identification

For the purpose of estimating the effect of the DST policy on energy use, a fundamental difficulty is that one cannot simultaneously observe both, how a state consumes energy under DST (the treatment) and how this state would behave in the absence of the treatment (the counterfactual). The optimal experiment would be to randomly allocate different timing schemes across states. While such an experiment cannot be observed, we believe that the DST modification that occurred in Australia in 2000 comes close. In this case we directly benefit from observations during the treatment period and the control period in both the treated and the non-treated state.

While we noted that the DST extension was implemented solely for operational purposes, and that we are not aware of any energy-based justifications, there may still be reasons to suspect that electricity consumption may have changed significantly even absent a DST extension. The 2000 Games were the most heavily visited Olympics event in history, school vacations were rescheduled to facilitate participation in carnival events, and the Games were watched on public mega screens and private TVs by millions of Australians in Sydney and elsewhere.

Our identification strategy incorporates several features designed to account for these potential confounders. First, we exclude the seventeen days of the Olympic period from the definition of the treatment period—this allows us to avoid many of the biases noted above. Second, even with the Olympics excluded from the treatment, electricity demand may have been affected before and after the games—for example by pre-Olympic construction activities and by extended tourism. To control for these, we ignore NSW (where the Olympics took place), and focus on the change in electricity demand in VIC relative to that in SA. This technique eliminates the impact of any confounders that operate on a national level.[10]

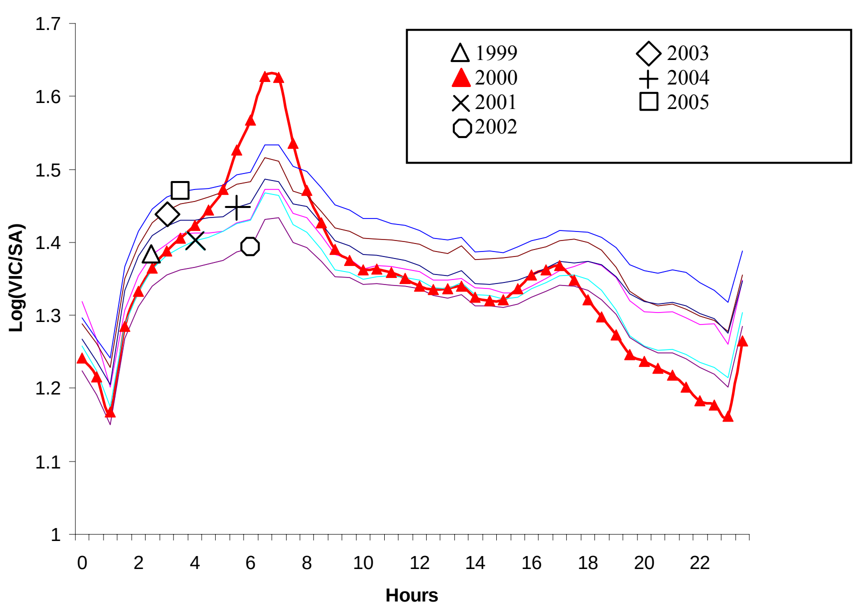
To control for unobservables that may have affected VIC and SA differentially, we use relative demand in the afternoon as an additional control. That is, because DST does not affect demand in the middle of the day, variations in demand levels that are not explained by observables such as weather can be attributed to non-DST-related confounders. With that, our model is robust against any "level shocks" affecting the level of the consumption in any state at any day *d*, but do not affect the shape of the half-hourly load pattern at date *d*. We verify the assumption that DST

[10] To further analyze whether visitors before and after the Olympic Games spent extended vacations in VIC or SA, we collected tourism information. These data clearly show that while NSW was affected by tourism in September, VIC and SA were unaffected. Details on the tourism data are provided in Appendix D.

does not affect afternoon demand by examining time changes in non-treatment years, as described in appendix E.

These three features of our model imply that a mild identifying assumption guarantees that our regressions produce an unbiased estimate of the extension's effect. We assume that, conditional on the observables and in the absence of the treatment, the ratio of VIC demand to SA demand in 2000 would have exhibited the same half-hourly pattern (but not necessarily the same level) as observed in other years. Strong support for this is found by plotting the ratio of consumption in VIC to that in SA for 1999-2005, as shown in Figure 5. The demand ratio exhibits a regular pattern in all non-treated years, even without controlling for observables. The figure also illustrates the large intra-day shift in consumption that occurred in VIC in 2000, in response to the DST extension.

Figure 5: Demand ratio between VIC (treated) and SA (control) averaged between 27 August and 27 October



Moreover, the level of the log ratio is unsystematically changing from smallest to largest over the years 2002, 2000, 2001, 1999, 2004, 2003, to 2005. This is consistent with the premise that level shocks, which we control for, affect one or the other state temporarily only. Finally, the decrease in evening demand in VIC in 2000 and the increase in the morning are clearly visible, being consistent with the analysis of section 3.

4.4.2 Treatment effect model

Our specification of the treatment effect model is primarily drawn from the difference-in-differences (DID) literature (Meyer 1995 and Bertrand et al., 2004). We augment the standard DID model by estimating a "triple-DID" specification because, as outlined in section 4.1, our control structure is three-fold:

(a) cross-sectional over states (using VIC as the treated state and SA as the control)

(b) temporal over years (using the untreated years in SA and VIC as controls)

(c) temporal within days (using afternoon hours as "within-day" controls)

Our linear specification is

$$\ln(q_{idh}) - \ln(\overline{q}_{id}) = T_{idh}\beta_h + X_{idh}\alpha_h + W_{idh}\phi_h + \varepsilon_{idh} \tag{1}$$

The dependent variable for each observation is the difference in logs between demand, q, in state i in day d in half-hour h, and $\overline{q}$, the average electricity demand between 12:00-14:30 in the same state and day, whereby $i \in$ {VIC, SA}, d = {1,2,…,186}, and h = {1,2,…,48}. The reference case model uses data from VIC and SA during 27 August to 27 October in 1999, 2000, and 2001; that corresponds to the time period of the DST extension in 2000, while in the years 1999 and 2001 during this period Standard Time is observed.

The covariates of primary interest are the indicator variables T_{idh} for the treatment period, unpooled by half-hour and active from 27 August to 14 September, 2000 and from 2 October to 28 October, 2000, thereby excluding the Olympic Games from the treatment.

Dummy variables X_{idh} include 48 half-hour dummies, and interactions of these dummies with indicator variables for the following: state, year, day of week, holidays, school vacations, transition days, the interaction of state with week of year, and the interaction of state with a flag for the Olympic period. The weather variables W_{idh} are also interacted with half-hour dummies[11] and include a quadratic in hourly heating degrees,[12] daily sunlight hours, the interaction of sunlight with temperature, hourly precipitation, the interaction of precipitation with temperature, and the average of the afternoon heating degrees. All weather variables enter the model lagged by one hour.

In equation (1) the treatment effect parameters to be estimated are given by β_h. The percentage change in electricity demand caused by the DST extension is given by

[11] Our final specification pools some hours to improve efficiency of the weather models. Doing so has no impact on the reported results on the treatment effects. In total, this specification has 48 treatment effects, 1019 fixed effects, 288 variables characterizing different days of the week, 144 variables to account for school-vacations, holidays, and transition days, 222 weather related variables and 96 indicators to dummy out the Olympic period.

[12] Heating degrees are calculated as the difference between the observed temperature and 18.33^{o} Celsius (65^{o} Fahrenheit). The motivation of squaring the heating degree is that as the temperature deviates from 18.33, cooling or heating efforts increase nonlinearly. This functional form is consistent with other electricity demand models in the literature (see Bushnell and Mansur 2005).

$\exp(\beta_h)$-1.[13] The main parameter of interest, however, is the percentage change in demand aggregated over all 48 half-hours, given by

$$\theta = \frac{\sum_{h=1}^{48} \exp(\beta_h)\omega_h}{\sum_{h=1}^{48} \omega_h} - 1. \qquad (2)$$

That is, θ is calculated as the weighted sum of the half-hourly percentage effects, where the weights ω_h are the average of the baseline 1999 and 2001 half-hourly demands during 27 August to 27 October, exclusive of the Olympic dates.

Our objective is to obtain the mean and other statistics of interest of the probability density function of the estimate $\hat{\theta}$, denoted $g(\hat{\theta})$. Because $\hat{\theta}$ is the weighted sum of non-iid lognormal variables, this distribution does not have a closed form solution and must be estimated numerically.[14]

To do so, we first develop a covariance estimator for $\hat{\boldsymbol{\gamma}} = [\hat{\boldsymbol{\beta}}^{\mathsf{T}}\ \hat{\boldsymbol{\alpha}}^{\mathsf{T}}\ \hat{\boldsymbol{\varphi}}^{\mathsf{T}}]^{\mathsf{T}}$, which in turn relies on the covariance structure of the disturbance $\boldsymbol{\varepsilon} = \mathbf{Y}\text{-}\mathbf{Z}\boldsymbol{\gamma}$. We allow $\boldsymbol{\varepsilon}$ to be both heteroskedastic and clustered on a daily level,

$$E(\varepsilon_{idh}\varepsilon_{idh}|\mathbf{Z}) = \sigma^2_{idh}, \qquad E(\varepsilon_{dj}\varepsilon_{dk}|\mathbf{Z}) = \rho_{dj}\ \forall\ j{\neq}k, \qquad E(\varepsilon_d \varepsilon_{d'}^{\mathsf{T}} \mid \mathbf{Z}) = \mathbf{0}\ \forall\ d{\neq}d'.$$

The motivation for selecting this block-diagonal structure is that it accounts for autocorrelation as well as for common shocks that affect both states

[13] To derive $\exp(\beta_h)$, we make use of the afternoon assumption that $E[\ \overline{q}_{id}\ |T_{id}{=}1] / E[\ \overline{q}_{id}\ |T_{id}{=}0] = 1$.

[14] Dependence between the estimates of the neighboring half-hours, $\hat{\beta}_h$ and $\hat{\beta}_{h-1}$ theoretically can lead to an a-typical shaped distribution g (see e.g. Vanduffel, 2005 for a recent treatment). Dependence structures vary by different covariance estimators. This is further illustrated in Appendix E.

contemporaneously. The clustered sample covariance matrix estimator is therefore used for γ (Wooldridge, 2003; Bertrand et al., 2004).

As an alternative to the clustered disturbance structure, we also estimate the model using the Newey and West (1987) estimator with 50 lags.[15] To adapt this estimator to our panel data, we block-diagonally partition the covariance matrix of ε into six groups (the three years by two states) and do not permit the lag structure to overlap across groups. For each block $\mathbf{\Omega}_j$, j=1,..,6, we assume the same covariance so that $\mathbf{\Omega}_j = \mathbf{\Omega}$.

With an estimate of the covariance of $\hat{\boldsymbol{\beta}}$ in hand, we numerically estimate the probability distribution $g(\hat{\theta})$ by taking 100,000 draws from the distribution $N(\hat{\boldsymbol{\beta}}, \mathbf{Cov}(\hat{\boldsymbol{\beta}}))$, and calculating $\hat{\theta}$ by (1) for each draw. It turns out that this numerical estimation produces a distribution $\hat{g}(\hat{\theta} \mid \mathbf{Z})$[16] that is indistinguishable from a normal distribution with a mean given by the empirical analogue of (2),

$$\hat{\theta} = \frac{\sum_{h=1}^{48} \exp(\hat{\beta}_h)\omega_h}{\sum_{h=1}^{48} \omega_h} - 1, \tag{3}$$

and a variance $\hat{\theta}$ calculated by the delta method,

$$V(\hat{\theta}) = \nabla_{\beta}\theta(\hat{\boldsymbol{\beta}})^{\mathrm{T}} \mathbf{Cov}(\hat{\boldsymbol{\beta}}) \nabla_{\beta}\theta(\hat{\boldsymbol{\beta}}), \tag{4}$$

---

[15] 50 lags allow the errors to be correlated over slightly more than one full day. Tests of AR(p) models on ε suggest that the disturbances are correlated over the first six hours of lags, but not beyond that. However, the coefficient on the 48th lag is significant. Also, note that the triple DID specification considerably decreases the autocorrelation properties of the dependent variable, relative to a standard DID. See Bertrand et al., 2003 for a discussion of the problems of autocorrelation and DID models.

[16] Appendix E compares the numerical with the analytical approximation methods. The 'hat' on g indicates that this distribution is itself estimated using the numerical approximation. Strictly speaking, we estimate the posterior of $\hat{\theta}$ that is conditional on $\mathbf{Z}$.

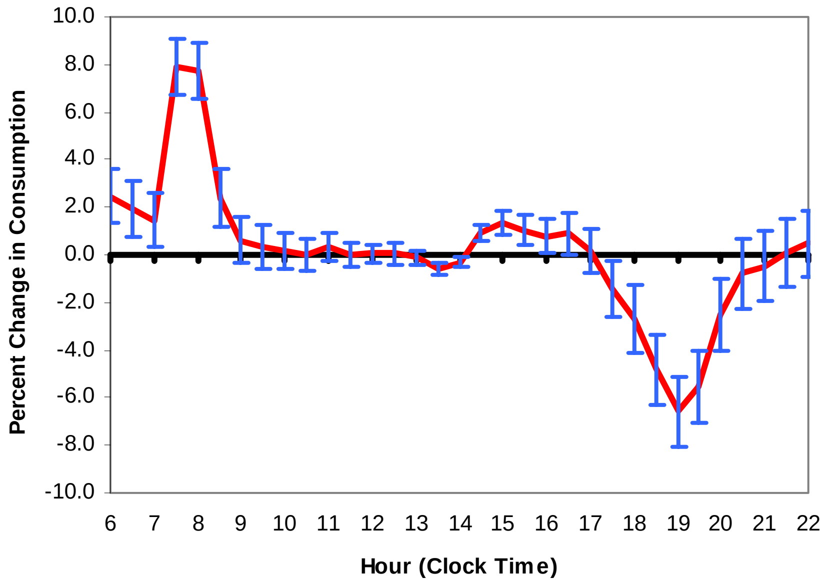
with $\nabla_{\beta}\theta$ as the (48 x 1) gradient vector of $\theta(\cdot)$ evaluated at $\hat{\boldsymbol{\beta}}$. We therefore report $\hat{\theta}$ and $V(\hat{\theta})$ as estimated by (3) and (4), rather than as the mean and variance of $g(\hat{\theta})$ and we can directly approximate any further statistic used in the below hypothesis tests as a Student's *t* distribution, which leads to the same results as if one were bootstrapping throughout.

4.5 Results

4.5.1 Reference case results

The primary goal of the empirical analysis is to examine the effect of the two-month extension of DST on electricity consumption. Figure 6 displays the estimated percentage impact of the DST extension on electricity demand in each half hour; these are the point estimates given by $\exp(\hat{\beta}_h)-1$. Extending DST affects electricity consumption in a manner consistent with the preliminary graphical analysis: there is a transfer in consumption from the evening to the morning. This behavior agrees with the expected effects of DST's one-hour time shift. Less lighting and heating are required in the evening; however, demand increases in the morning—particularly from 07:00 to 08:00—driven by reduced sunlight and lower temperatures.

Figure 6: Half hourly treatment effects of extending DST on electricity use



The estimated effect of extending DST in VIC, disaggregated by half-hour, with 95% confidence intervals. Standard errors are clustered by day.

To assess whether the evening decrease in demand outweighs the morning increase, we aggregate the half-hourly estimates using (3) to yield an estimate of θ. We find that the extension of DST failed to conserve electricity. The point estimate of the percentage change in demand over the entire treatment period is +0.11% with a clustered standard error of 0.39.[17]

[17] In DID panel settings, Bertrand et al. caution that results are sensitive with respect the chosen standard errors. Our results very clearly confirm such bias. In our case, assuming homoskedasticity would result in a standard error of $\hat{\theta}$ of 0.08. Instead applying the Newey-West covariance estimator results in a standard error of 0.32. Although the Newey-West correction in large sample sizes promises a good approximation, here we chose to report our main results using the more conservative clustered standard errors (0.39). For a discussion on the comparison between the latter two approaches, see Petersen, 2006.

Table 3: Summary of percentage change treatment effects

	All days	"September"	"October"	Weekdays	Weekends
% change	0.11	0.34	-0.06	0.44	-1.94
standard error	(0.39) [0.32]	(0.43) [0.34]	(0.43) [0.36]	(0.40) [0.33]	(0.41) [0.40]

Clustered standard errors are in parentheses and Newey-West standard errors are in brackets.

We also examine the impact of the DST extension separately for the "September" period and the "October" period.[18] Because September in the southern hemisphere is seasonally equivalent to March in the northern hemisphere, this examination has policy implications—recent efforts to extend DST in the U.S., California, and Canada concern an extension into March, as DST is already observed in April in these locations. Prior studies suggest that such an extension creates electricity savings of 1% (U.S.), 0.6% (California), and 2.2% (Ontario, Canada). By contrast, our estimate shows that the extension of DST into September in Australia *increased* electricity demand by 0.34%.[19] This result raises a concern that extending DST in North America will fail to yield the anticipated electricity savings.

To formally compare our estimates to the previous literature, we define four null hypotheses, H_0:, (1) $\theta = 2.2\%$, (2) $\theta = 1.0\%$, (3) $\theta = 0.6\%$, and (4) $\theta = 0.0\%$. In each case, the alternative, H_A, is that the change in electricity demand is greater than the cited value. Table 4 displays *p*-values for rejection of each null hypothesis, given both our overall estimate and our unpooled estimate. Even with conservative clustered standard errors, we reject at the 5% level the most modest estimate of the prior

[18] "September" covers the time period from 27 August, 4am to 14 September, and "October" covers 2 October to 27 October—these dates correspond to the treatment period in 2000: the extension period excluding the 17 days of the Olympic games.

[19] The point estimate in "October" is that the extension conserves electricity by 0.06%. While the difference between the "September" and "October" estimates is significant at only the 30% level, the sign of the difference is intuitive: in "October" there is more morning sunlight and temperatures are warmer, so the morning increase in demand is mitigated.

literature—a 0.6% reduction in electricity use in September. Over the entire treatment period, we reject a 1% reduction in demand at the 1% level, and reject a 0.6% reduction at a 10% level. These rejections are strengthened with the use of Newey-West standard errors.

All told, our results indicate that claims that extending DST will significantly decrease energy use and GHG emissions are at best overstated, and at worst carry the wrong sign. In particular, a long, two-month, extension is more likely than not to increase electricity consumption.

Table 4: *p*-values of testing the energy saving hypotheses

		"September" ($\hat{\theta}$ =+0.34%)	**"September" and "October"** ($\hat{\theta}$ =+0.11%)		
	Null hypothesis	Cluster	Cluster	Newey-West	"OLS"
Electricity Savings	**-2.2%**	0.000***	0.000***	0.000***	0.000***
	-1%	0.003***	0.007***	0.001***	0.000***
	-0.6%	0.037**	0.075*	0.033**	0.000***
Electricity Neutrality	**0.0%**	0.292	0.384	0.375	0.135

*** rejected at p = 0.01, ** rejected at p = 0.05, * rejected at p = 0.1

4.5.2 Robustness

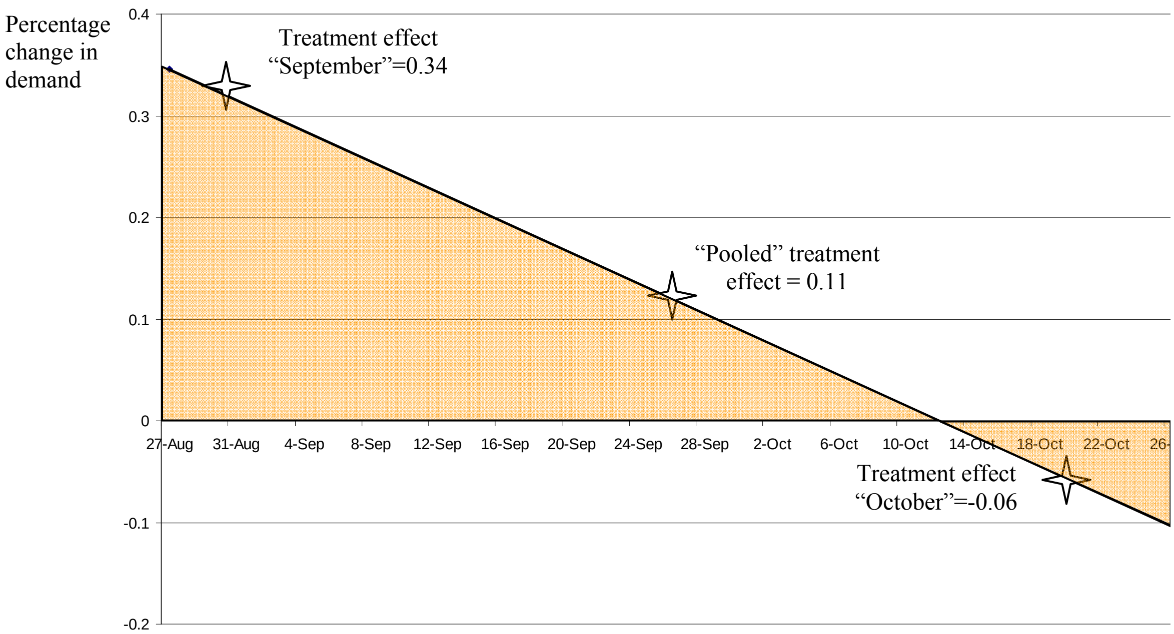
Our results are robust to many alternative specifications. The use of time trends rather than weekly dummies does not affect the results, nor do alternative weather specifications. In particular, our results are invariant to the choice between a weather model taken from Bushnell and Mansur (2005) and one from CEC (2001) (described in detail in section 7). Further, our results do not change if we include years and months of data beyond what we use in our reference case. This robustness is underlined by the precise fit of our model—the adjusted R^2 is greater than 0.94.

Regression equation (1) contains over 1800 parameters. While the point estimates and the standard errors for the parameters of primary interest—the treatment effect—are discussed above, most of the other coefficients are significant and carry signs that agree with intuition. For example, weekends, holidays, and vacations lower electricity consumption at all hours of the day and particularly in the morning. Deviations from the base temperature of 18 degrees Celsius increase electricity consumption, consistent with the effects of air-conditioning (when above 18 degrees) and heating (when below 18 degrees).

The weights ω_h used to calculate $\hat{\theta}$ are based on the average of the 1999 and 2001 half-hourly demands. As an alternative set of weights, we also use the estimated half-hourly counterfactual demand in 2000, given by $\exp\{X_{VICdh}\alpha_{VICdh} + W_{VICdh}\phi_{ih}\} \cdot \overline{q}_{VICd}$. Doing so does not affect our estimate of $\hat{\theta}$.

To verify the robustness of our unpooled result, we modify the pooled specification to include the interaction of the treatment dummies with a daily time trend. That is, we add the term $t \cdot T_{idh} \cdot \beta_h^t$ to regression specification (1) for each half-hour $h = 1,\ldots,48$, where t denotes the day of the year. Figure 7 displays the estimated treatment effect over the period 27 August to 27 October (calculated as $\theta(t) = [\Sigma_h \exp\{\beta_h + t \cdot \beta_h^t\} \cdot \omega_h / \Sigma_h \omega_h)]-1)$. Victoria marginally benefits from DST after 14 October; however, DST increases energy use prior to this date. This result agrees with our unpooled "September" and "October" treatment effects.

Figure 7: Optimal timing of DST



As a final check of our estimates, we evaluate whether extending DST causes relatively greater reductions in electricity consumption on weekends and holidays than on working days. This would be consistent with the intuition that, on non-working days, less early activity will mitigate the morning increase in demand. We estimate that electricity consumption on working days increased by 0.4% during the extension, while consumption on weekends and holidays decreased by 0.9%. This difference is significant at the 2% level.

4.6 Alternative methods to measure the effect of DST on electricity use

In the remainder of the paper we examine two alternatives to measure the effect of DST on energy. This is useful for at least two reasons: first, the Australian data provide us with the unique opportunity to evaluate the proposals to extend DST in the U.S., Canada, New Zealand and Australia (*Energy Policy Act*, 2005; *Joint Senate Resolution*, 2001; Young, 2005; Eckhoff, 2001; Hansard, 2005) as we can analyze the predictive power of the prior modeling approaches. Second, the data provide a validation tool to examine the structure of the prior modeling methods of the DST literature, which can be categorized into two types: the "week before / week after" technique (Eckhoff, 2001; Young, 2005) and the simulation approach (Rock, 1997; CEC, 2001).

The simulation approach uses data on hourly electricity consumption under the status quo DST timing policy to simulate consumption under a DST extension. This procedure first employs a regression analysis to assess how electricity demand in each hour is affected by light and weather, and then uses the regression coefficients to predict demand in the event of a one-hour time shift lagging the weather and light variables appropriately. The simulation results rely on the assumption that extending DST will not cause new patterns of activity than those observed in the status quo. This may not hold in practice. For example, to simulate demand under extended DST at 07:00, the model must rely on observed status quo behavior at 07:00 under cold and low-light conditions. Without a DST extension, these conditions are observed only in mid-winter. The simulation will be inaccurate if people awaken later in winter than they do in spring under extended DST, perhaps because they rise earlier as they become accustomed to increasing morning light in the spring and continue this behavior even after the extension causes mornings to be dark again.

With the Australian quasi-experiment, by contrast, we can estimate the treatment effect directly, based on the comparison of both regimes, the *status quo* and the *treatment period* (the period of the DST extension in 2000). By re-estimating the simulation models based on the status quo observations and then forecasting the electricity demand under the treatment, we have a tool to evaluate the performance of this approach in detail.

The "week before / week after" technique examines electricity use before and after the existing spring and fall time changes. These studies confirm the conventional wisdom that DST saves energy. However, an extension introduces DST to a time of year when the days are shorter and cooler than they are when the time shift usually occurs. Secondly, the first week of DST has longer and warmer days than the week prior to the springtime change. Therefore, these studies likely overestimate the energy savings of an extension.

We first show that both methods significantly overstate electricity savings. We then try to understand why these biases arise. We find that by carefully modifying the sample selection, the simulation models' aggregate predictive power improves; however, they still fail to accurately predict the intraday changes in demand. For the "week before / week after technique", we show that by controlling for differences in weather reduces the bias, but the variance of the estimates remains high. Overall, these results cast suspicion on the models' previous policy applications.

4.7 Evaluation of the Simulation Approach

A natural question to ask is whether or not the simulation approach would have predicted the DST effects sufficiently well. To test the simulation approach, we employ the most recent model developed by CEC, 2001, which has been used in the U.S. to argue in favor of a year-round DST extension in California. The first stage of the model is a regression of hourly electricity demand, q_{dh}, on employment, weather, and sunlight variables:

$$q_{dh}^{sim} = a_h + b_h \text{Employment}_{dh} + c_h \text{Weather}_{dh} + d_h \text{Light}_{dh} + u_{dh}$$

The disturbance $\boldsymbol{u}_d$ is correlated across the h = 1,…,24 hourly equations, per the Seemingly Unrelated Regression method (Zellner, 1962). The regression allows the weather and light coefficients to vary across the twenty-four hours of the day, and the weather specifications are very detailed. For example, the temperature variables are separated into hot, cold, and warm days, because a hot hour which follows other hot hours will have higher electricity demand than a hot hour which follows cool hours (because buildings retain heat).[20] Once the vector of regression coefficients is

[20] For each half hour the weather and light regressors consist of temperature variables by (1) a one-hour weighted average of its quadratic and cubic, where the weights are .45 times the temperature in the hour that includes the last half-hour of an electricity use hour, .45 times the temperature in the hour that includes the first half-hour of an electricity use hour, and .10 times the previous hour; and (2) a three day weighted average of the temperature separately for hot spells, warm spells and cold spells, with 60% weight on average temperature one day lagged, 30% on 2 days lagged, and 10% on 3 days lagged. Hot, warm and cold are defined by the temperature cut-off

estimated, they are used in the second stage to forecast electricity consumption under a DST extension. This is accomplished by lagging the weather and lighting variables by one hour and adding the first stage realized error term to project

$$\hat{q}_{dh}^{sim} = \hat{a}_h + \hat{b}_h \text{Employment}_{dh} + \hat{c}_h \text{Weather}_{dh-1} + \hat{d}_h \text{Light}_{dh-1} + \hat{u}_{dh} \quad \forall d \in \{\underline{D}, \underline{D}+1, ..., \overline{D}\}$$

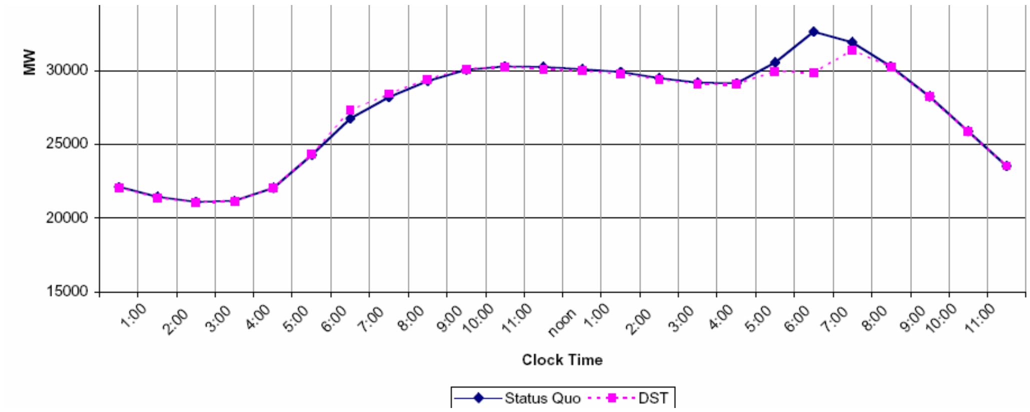
for the days $d = \underline{D}, ..., \overline{D}$ for which a DST extension is being considered.

Figure 8 displays observed electricity demand in California during March 1998-2000 when Standard Time was in effect, as well as the simulated demand for extended DST. Recall that March in California is equivalent to September in Australia. The simulation predicts that under DST electricity consumption will be significantly lower in the evening, between 17:00 and 19:00, leading to an overall 0.6% decrease in electricity use for the month of March.

---

values 21.11°C and 10.00°C. Humidity, precipitation, barometric pressure, wind speed, visibility, and cloud cover also enter the weather specification. The lighting variables are the percentage of the hour in daylight throughout California and the percentage in twilight. The light variables are included only for those hours in which light conditions vary over the year, under either standard time or DST. Details on the definition on these variables, the estimation of the model and simulation are explained in CEC, 2001.

Figure 8: If DST had been imposed in March 1998-2000 in California



Source: CEC, 2001.Actual status quo demand is observed under Standard Time. The forecasted demand is simulated under the assumption that DST had been imposed. For California, the observed and simulated load shapes for a DST extension into January and February look similar. More details are provided in CEC, 2001.

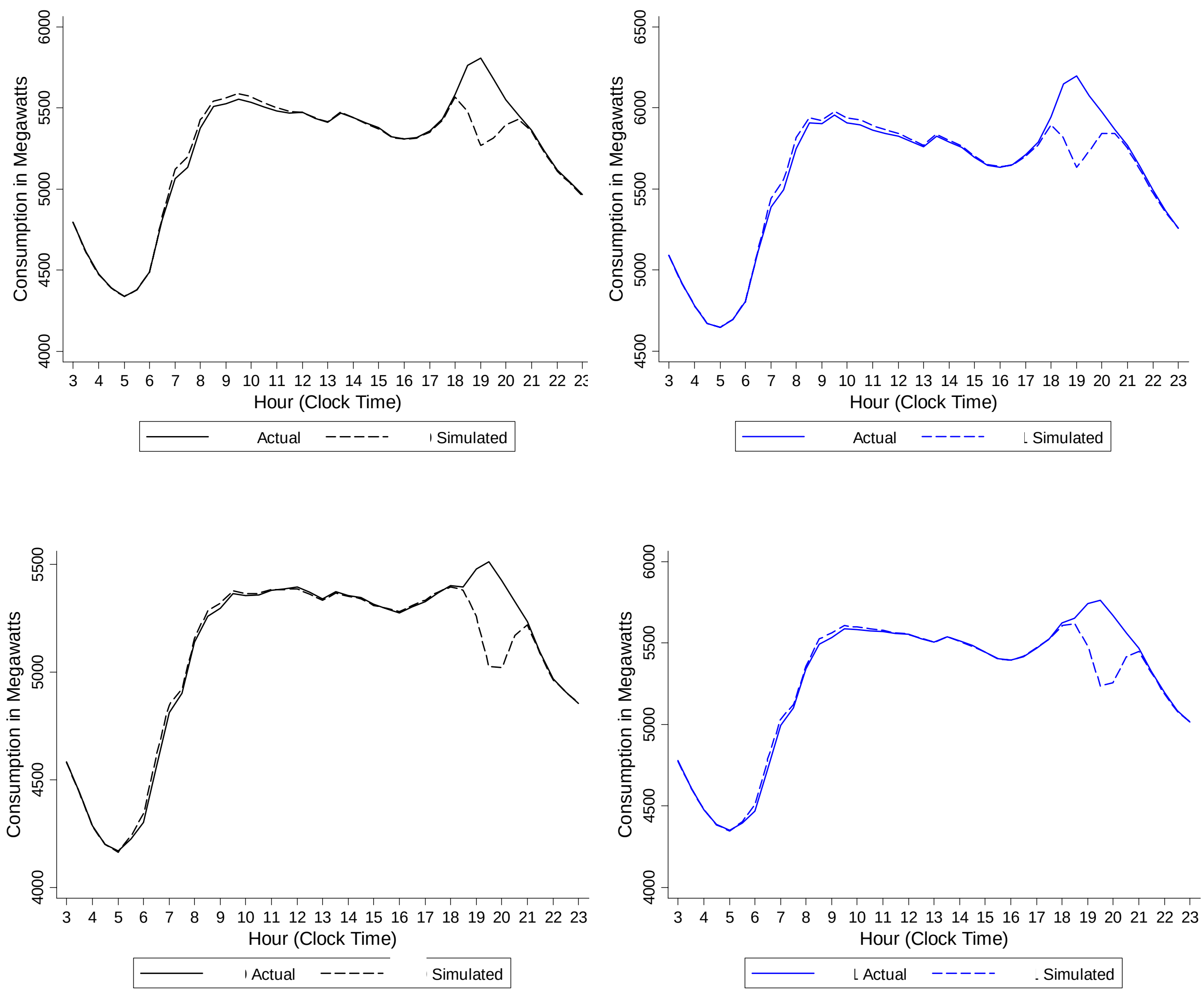
We apply the CEC model to the Australian data for the state of VIC, with a few changes to the specification.[21] Figure 9 illustrates the simulated electricity demand under a DST extension in "September" and "October". The simulated load shapes in VIC very closely resemble those for the California simulation, and predict energy savings of 0.41% to 0.44%.

Figure 10 compares the characteristics of actual demand under the VIC treatment with simulated consumption. The figure shows that the simulation fails to predict a morning increase in electricity consumption similar to that observed in 2000, and also overestimates the evening savings. The simulated decrease in consumption is

[21] Instead of using 24 hourly equations, we take advantage of the more detailed Australian dataset and estimate the model with 48 half-hourly equations. We also improve the explanatory power of the model by including six day-of-week dummies and an indicator variable for vacations, holidays, and transition days. Finally, the Australian weather data do not contain variables for visibility and cloud cover that were used in CEC, 2001. Instead we use the number of hours of sunshine per day and the interaction of this variable with temperature. Also, the humidity and precipitation variables are correlated with visibility. In total the model applied to Australia has 1052 parameters to be estimated (48 equations with 24 parameters each) based on the data from 1 January, 1999 to 31 December, 2002, but excluding the treatment period in 2000.

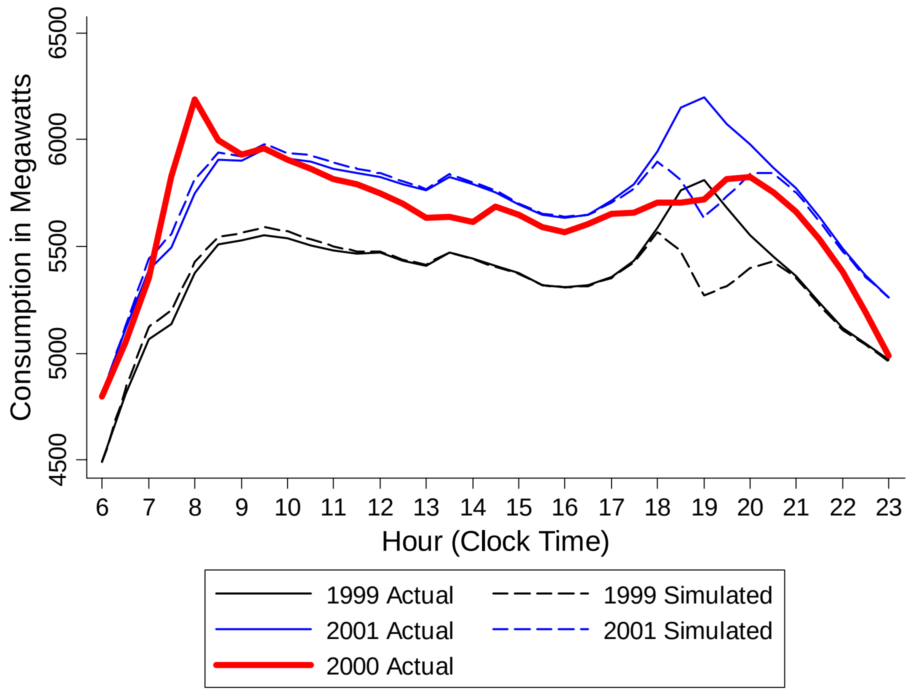
inconsistent with what actually happened in VIC. Based upon our triple DID estimate and clustered standard error presented earlier, we reject the -0.41% prediction of the simulation at a 5% significance level.

Figure 9: Actual vs. forecasted VIC demand based on the CEC simulator



Actual demand is observed under Standard Time. The forecasted demand is simulated under the assumption that DST had been imposed.

Figure 10: Actual and simulated electricity consumption in VIC over "September" in various years. DST is in effect only during 2000



Average electricity consumption in VIC by half-hour in "September" in various years. Solid lines represent observed consumption, and dashed lines represent simulations of what consumption would have been if DST were observed.

The first row of Table 5 summarizes our simulation results. It is striking that for all the periods from 1999 to 2001, the estimates of energy savings fall in a narrow range from 0.41% to 0.45% and strongly reject our treatment effect estimate of section 6.[22] Table 5 further displays the test statistics for the comparison of the simulation results to a 0.6% reduction in energy use—the simulated prediction for California (CEC, 2001). Our simulations cannot reject savings of 0.6%, confirming the preliminary result that the VIC simulation is very similar to that for California.. As a

[22] To perform the hypothesis tests we need to calculate the variance of the sum of simulated energy demand $\Sigma_{d=\underline{D}}^{\overline{D}}\Sigma_{h=1}^{48} q_{dh}^{sim}$. This is given by $\Sigma_i\Sigma_j[\mathbf{X}_{\text{sim}}{}^{T}\mathbf{Cov(\beta)X}_{\text{sim}}]_{ij}$, that is as the sum of the elements of the matrix $\mathbf{X}_{\text{sim}}{}^{T}\mathbf{Cov(\beta)X}_{\text{sim}}$, whereby $\mathbf{X}_{\text{sim}}$ is the block-diagonal "simulation" regressor matrix of dimension $48{\cdot}(\overline{D} - \underline{D})$ x 1052 with each block h = 1,2,…,48 defined as columns of [1, Employment$_{dh}$, Weather$_{dh\text{-}2}$, Light$_{dh\text{-}2}$, Weekday1$_{dh}$,…,Weekday6$_{dh}$, Workday$_{dh}$,] and **Cov(β)** is the 1052 x 1052 estimated covariance matrix of **β**.

robustness check we repeat this exercise for the month of October (which is equivalent to the month of April in the northern hemisphere), leading to very similar results.

Table 5: Simulating a DST extension using the CEC methodology

Year		1999		2001		"September" 1999	"September" 2001
Period		September	October	September	October		
%-change between DST and Standard Time		-0.44	-0.44	-0.43	-0.41	-0.43	-0.41
t-value with respect to	energy neutrality	-2.02	-1.82	-1.42	-1.64	-1.81	-1.40
	energy savings of 0.6%	0.72	0.64	0.54	0.53	0.73	0.65

We attempted to understand the causes of the simulation's misprediction. We found that, by shrinking the sample in the first stage regression, the predictive power can be increased considerably.[23] We use a sample period in which sunset, sunrise, light and weather conditions are most similar to the simulated extension period in September.[24] Table 6 displays the regression results from the revised simulation model—the results now show that the DST impacts are statistically indistinguishable from zero, which more closely corresponds to what actually happened in VIC. Also, with this improved specification the prior electricity savings estimates of 0.6% and 1% in the U.S. are now rejected at the 10% significance level and lower. However, when we analyze the refurbished model on a half-hourly basis we still find that it substantially under-predicts morning electricity demand between 07:00 and 09:00, and

[23] The original simulation models' parameters are estimated based on the status quo data from all twelve months of the year. On the one hand, one might expect that this variation in weather improves the forty-eight weather models especially because they explicitly account for the nonlinearities and discontinuities by use of hot, warm and cold weather spells. On the other hand, we show that significant improvements are made by being more selective.

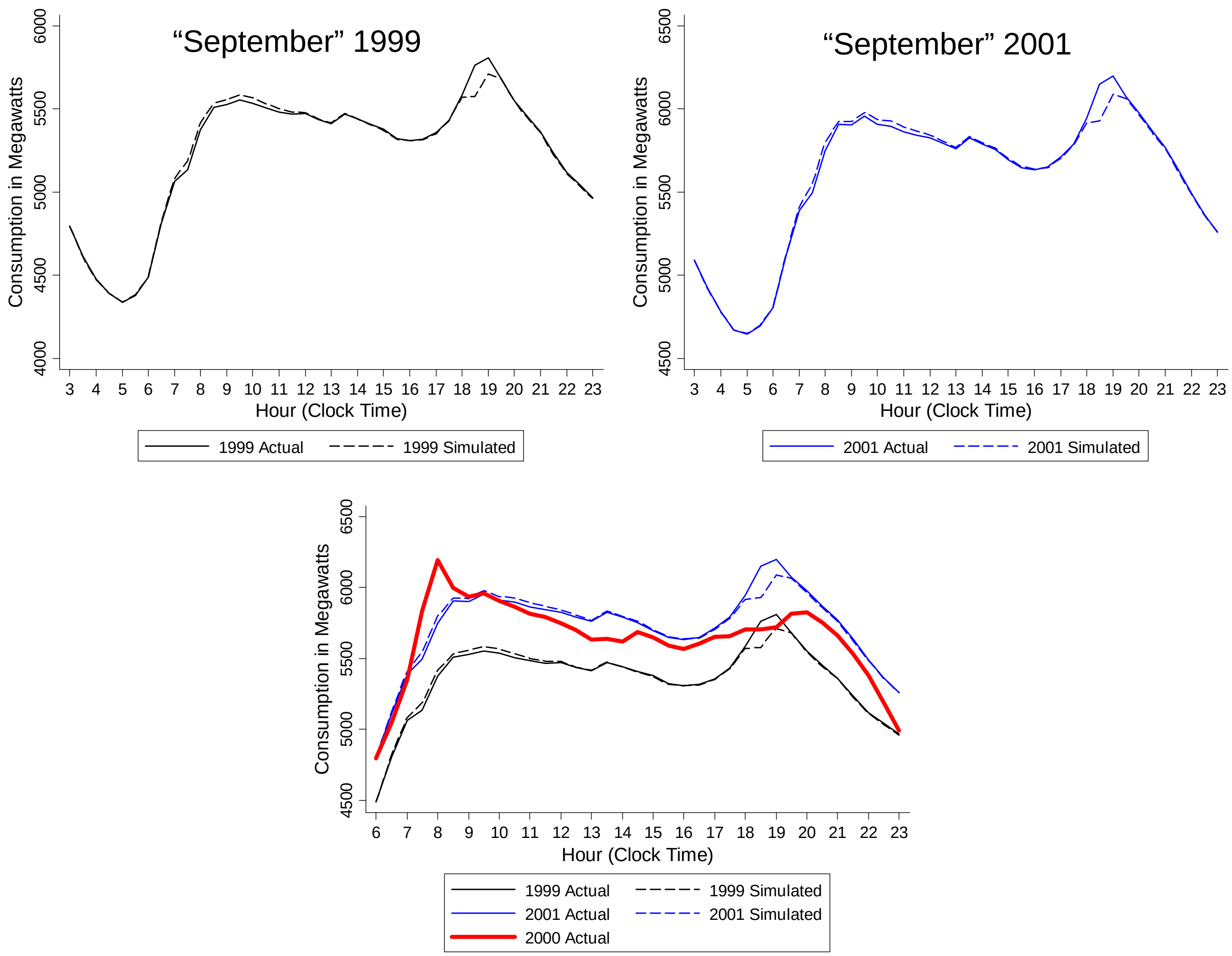
[24] For example, to predict an extension into September, we suggest to limiting the sample size to the months from March to September and excluding the full month of July and the first half of August. See the sunrise, sunset, weather table A1 in Appendix A for the more detailed motivation for choosing these periods.

over-estimates the evening demand. These two mispredictions cancel one another, leading to the more accurately predicted overall effect. We conclude that despite extensive adjustments this simulation model cannot predict the substantial intra-day shifts that occur due to the early adoption of DST.

Table 6: Simulating a DST extension using the refurbished simulation model

Year		1999	1999	2001	2001
Period		September	"September"	September	"September"
%-change between DST and Standard Time		-0.005	-0.027	-0.026	-0.025
t-value with respect to	energy neutrality 0.0%	-0.02	-0.01	-0.06	-0.07
	energy savings 0.5%	1.75	1.6	1.210	1.24
	energy savings 0.6%	2.10	1.92	1.34	1.50
	energy savings 1%	3.51	3.19	2.27	2.54

Figure 11: Actual versus simulated VIC demand based on the refurbished simulator



4.8 Evaluation of the "week before / week after technique"

Applying the "week before / week after technique" (WBAT), to VIC, as used in New Zealand and Canada, would lead to a prediction of electricity savings of 1.77%.[25] The point estimate is slightly lower than the savings predicted in Ontario (2.2%) and New Zealand (2.0%-3.5%), however, with a clustered standard error of 1.60, the estimate is statistically insignificant. Still, this correlation estimate is consistent with the intuition that in summertime the requirement for indoor electricity use decreases due to improved weather conditions. Once we control for weather and day weekday/workday dummies, however, the point estimate on the DST coefficient increase to +1.11%. Table 7 shows that varying the number of days before and after the springtime change causes the WBAT estimates to vary from -1.21% to -1.77% when weather variables are not included, and from 0.52% to 1.11%, when weather variables are included. This variance is not surprising given the large standard errors. We believe that this lack of robustness makes this approach unsuitable for policy analysis in Australia.

Table 7: Percentage change due to DST using the "week before / week after technique"

	Days used before and after the springtime change					
	7 days		10 days		4 days	
	%-change	s.e.	%-change	s.e.	%-change	s.e.
WBAT	-1.77	1.60	-1.34	1.29	-1.21	1.92
WBAT conditional on weather	1.11	0.69	0.52	0.55	0.72	1.31

Standard errors (std) based on clustered covariance matrix by date. The original WBAT approach employs data of one week before and one week after the springtime changes over the years from 1999 to 2005, excluding the year 2000 (7 days column).

[25] The WBAT approach used data one week prior to and one week after the springtime changes over the years from 1999 to 2005, excluding the year 2000.

4.9 Summary and Conclusions

Given the economic and environmental imperatives driving efforts to reduce energy consumption, policy-makers are considering extending Daylight Saving Time (DST). Doing so is widely believed to reduce electricity use.[26] Our research challenges this belief, as well as the studies underlying it. We offer a new test of whether extending DST decreases energy consumption by evaluating an extension of DST that occurred in the state of Victoria, Australia in 2000. Using half-hourly panel data on electricity consumption and a triple-difference treatment effect model, we show that, while extending DST does reduce electricity consumption in the evening, the increased demand in the morning cancels these benefits out. We statistically reject electricity savings of 1% or greater at a 1% significance level.

We also cannot confirm two additional DST extension benefits that have been discussed in California: a reduction in electricity prices and a reduction in the likelihood of blackouts driven by a more balanced hourly load shape. We instead show that the Australian DST extension significantly increased expenditures on electricity and caused a sharp peak load in the morning.

From an applied policy perspective, this study is of immediate interest for Australia, which is actively considering an extension to DST. Moreover, the lessons from Australia may carry over to the U.S. and to California—Victoria's latitude and

---

[26] On signing the Energy Policy Act on 8 August, 2005, President Bush stated that it is primarily a "*security bill*" to become "*less dependent on foreign sources of energy*" (Bush, 2005). The U.S. government emphasized this by expressing the estimated 1% electricity savings of extended DST as "*to reduce energy consumption by the equivalent of 100,000 barrels of oil for each day of the extension*" (CENR, 2005).

climate are similar to those of central California.[27] In particular, the planned extension that will occur in the U.S. in 2007 will cause DST to be observed in March—a month that is analogous to September in Australia, when our point estimates suggest that DST will increase rather than decrease electricity consumption. With this, our results run contrary to recent simulation-based studies and suggest that current proposals to extend DST may be misguided.

To further investigate the relationship of our study to previous simulations, we re-estimate the simulation model that supported a DST extension in California, using Australian data. We find that simulation models over-estimate energy savings casting suspicion on its previous policy applications in the U.S. Similarly, we scrutinize the "week before / week after technique" which has been employed in Canada and New Zealand and find that this method also predicts savings that are too large.

It should be noted that our estimates of energy use are likely represent a lower bound, as we account for electricity consumption only. Considering gasoline demand as well may increase the estimate of DST's effect on energy consumption, as longer and warmer evening hours drive an increase in evening leisure travel (Lawson, 2001).

Finally, our study leaves scope for future work. First, an *ex-post* evaluation of the pending U.S. DST extension will be a worthwhile enterprise. Second, the non-energy impacts of extending DST also require investigation—potential studies include

[27] While we are not in a position to extend our results to *any* country, it is worth noting that there are several other major coastal cities around the world at approximately the same latitude as Melbourne (latitude 37.5 South)—for example, Buenos Aires (34.4) in the southern hemisphere and San Francisco (37.77), Washington D.C. (38.5) and Tokyo (35.4) in the northern hemisphere—locations within countries that are considering changes to their DST systems. These countries may find our results helpful in order to assess potential costs and benefits of such measures.

impact analyses on crime, traffic accidents, and economic coordination, which could build upon prior work in these areas (Coren, 1996; Coate and Markowitz, 2004; Kamstra et al., 2000; Lambe and Cummings, 2000; Varughese and Allen, 2001; Hamermesh et al., 2006). Such work will allow the research community to provide policy-makers with evidence to support informed decisions regarding the future status of DST

References

Adkins, L.C., D.S. Rickman and A. Hameed (2002): Bayesian Estimation of Regional Production for CGE Modeling. Paper presented at the fourteenth International Conference on Input-Output Techniques October 10-15, 2002, Montréal, Canada.

Aït-Sahalia, Y. and J. Duarte (2003): Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116, pp.9 – 47.

Aldrich, B. (2006): Daylight Saving Time, Its History and Why We Use It. California Energy Commission. Also downloadable at: http://www.energy.ca.gov/daylightsaving.html.

Andrews, D. (1999): Estimation When a Parameter Is on a Boundary. *Econometrica*, 67, 1341–1383.

Andrews, D. (2001): Testing When a Parameter Is on the Boundary of the Maintained Hypothesis. *Econometrica*, 69, 683–734.

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.

Apostolakis, B.E. (1990): Energy-Capital Substitutability / Complementarity: The Dichotomy. *Energy Economics*, 1:48-58.

Australian Bureau of Statistics (2001a): Tourism Indicators, Report 8634.0, December Quarter 2000, Canberra.

Australian Bureau of Statistics (2001b): Tourist accommodation: an analysis over the Olympic period. Tourism Indicators, December Quarter 2000.

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. 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.

Barnett, W.A., J. Geweke, and M. Wolfe (1991): Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model. *Journal of Econometrics*, 49, No. 1/2, pp.5-50.

Beauregard-Tellier, F. (2005): Daylight Saving Time and Energy Conservation. Economics Division**,** 29 July, 2005. Library of Parliament, Canada.

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.

Berndt, E.R. and M.S. Khaled (1979): Parametric Productivity Measurement and Choice among Flexible Functional Forms, *Journal of Political Economy*, **87**, 1220-1245.

Bertrand, M, E. Duflo, and S. Mullainathan (2004): How much should we trust differences-indifferences estimates? *Quarterly Journal of Economics,* 119, 249-275.

Blundell, R (2004): Presidential address of the EEA/ESEM (*European Economic Association* and *Econometric Society European Meeting*) conference in Madrid, August, 20, 2004.

Bush, G.W. (2005): Office of the Press Secretary, 8 August, 2005. President Signs Energy Policy Act, Sandia National Laboratory, Albuquerque, New Mexico.

Bushnell, J.B. and E.T. Mansur (2005): Consumption under Noisy Price Signals: A Study of Electricity Retail Rate Deregulation in San Diego. *Journal of Industrial Economics*, 53. pp. 493-513.

CEC, 2001: Effects of Daylight Saving Time on California Electricity Use. California Energy Commission. Report authored by Adrienne Kandel and Daryl Metz .

CENR (2005): Press release of the Committee on Energy and Natural Resources. Energy Policy Act of 2005 Bill Summary. Also available on the internet at: energy.senate.gov/public/_files/PostConferenceBill

Chambers (1988): Applied Production Analysis, The Dual Analysis. Cambridge University Press.

Chen, M.H., Q.M. Shao and J.G. Ibrahim (2000): Monte Carlo Methods in Bayesian Computation. Springer, New York.

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

Chichilnisky, G. and G.Heal (1983): Energy Capital Substitution: a General Equilibrium Analysis. IIASA, Laxenburg, Austria.

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

Coate, D. and S. Markowitz (2004): The effects of daylight and daylight saving time on US pedestrian fatalities and motor vehicle occupant fatalities. *Accident Analysis and Prevention*, 36, 351-357.

Coren, S. (1996): Daylight saving time and traffic accidents. *New England Journal of Medicine*, 334, 924.

Cuesta, R.A., C.J. O'Donnell, T.J. Coelli and S. Singh (2001): Imposing Curvature Conditions on a Production Frontier: With Applications to Indian Dairy Processing Plants. *CEPA Working Papers*, No. 2/2001, ISBN 1 86389 749 6, School of Economics, University of New England, Armidale.

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.

DOT (1975): The Daylight Saving Time Study: A Report to Congress by the US Department of Transportation. *Washington*, GPO, 1975. 2 v. HN49.D3U65 1975, Vol. 1, final report of the operation and effects of daylight saving time and Vol. 2, supporting studies: final report of the operation and effects of daylight saving time.

Downing, M. (2005): Spring Forward: The Annual Madness of Daylight Saving Time. Shoemaker Hoard, Washington D.C.

ECCJ (2006): Report on the National Conference on the Global Environment and Summer Time. The Energy Conservation Center, Japan (ECCJ). Available on the Internet at http://www.eccj.or.jp/SummerTime/conf/index_e.html

Eckhoff, G. (2001): Minister Urged to Consider Early Daylight Saving. Press Release by ACT New Zealand, published i.e. in *Scoop Independent News* on August 14, 2001.

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.*

EIA (2005): Direct Use and Retail Sales of Electricity to Ultimate Customers by Sector, by Provider. Report Released November 2005.

Emergency Daylight Savings Time Energy Conservation Act (1973): U.S. Public Law 93-182, H.R. 11324 (87 Stat. 707), signed by President Richard Nixon 15 December, 1973.

Energy Policy Act (2005): U.S. Public Law 109-58, signed into law by President George W. Bush, 8 August, 2005.

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

Filippini, M. and J. Wild (2001): Regional differences in electricity distribution costs and their consequences for yardstick regulation of access prices. Energy Economics, vol. 23(4), pp. 477-488.

Fischer, D., A.R. Fleissig and A. Serletis (2001): An Empirical Comparison of Flexible Demand System Functional Forms. *Journal of Applied Econometrics*, 16(1), pp.59-80.

Fleissig, A.R., T. Kastens and D. Terrell: (1997): Semi-nonparametric Estimates of Substitution Elasticities. *Economic Letters*, 54, pp.209-215.

Fleissig, A.R., T. Kastens and D. Terrell: (2000): Evaluating the Semi-nonparametric Fourier, AIM, and Neural Networks Cost Functions. *Economic Letters*, 68, pp.235-244.

Franklin, B. (1784): An Economical Project, Essay on Daylight Saving. Letter to the Editor. *The Journal of Paris*, April 26th, 1784.

Friesen, J. (1992): Testing Dynamic Specification of Factor Demand Equations for U.S. Manufacturing. *Review of Economics and Statistics*, Vol. 74,2,240-50.

Gallant, A. R. and G.H. Golub (1984): Imposing Curvature Restrictions on Flexible Functional Forms. *Journal of Econometrics*, 26, pp.295-322.

Geweke, J. (1986): Exact Inference in the Inequality Constrained Normal Linear Regression Model. *Journal of Applied Econometrics*, vol. 1, issue 2, pp. 127-41.

Griffiths, W.E. (2003): Bayesian Inference in the Seemingly Unrelated Regressions Model. In D.E.A. Giles, editor, Computer-Aided Econometrics, New York: Marcel Dekker, pp. 263-290.

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.

Griffiths, W.E., C.L. Skeels and D. Chotikapanich (2002): Sample Size Requirements for Estimation in SUR Models. In A. Ullah, A., Chaturvedi and A. Wan, Eds., Handbook of Applied Econometrics and Statistical Inference, New York: Marcel Dekker.

Hamermesh, D.S., Myers, C.K. and Pocock, M.L. (2006): Time Zones as Cues for Coordination: Latitude, Longitude and Letterman. National Bureau of Economic Research (NBER) Working Paper 12350.

Hansard (1999a): Legislative Assembly Hansard: Standard Time Amendment Bill, Second Reading, 26 May 1999, article 40, New South Wales.

Hansard (1999b): Legislative Council Hansard: Standard Time Amendment Bill, Second Reading, 2 June 1999, article 9, New South Wales.

Hansard (2005): Legislative Assembly Hansard: Standard Time Amendment (Daylight Saving) Bill, Article 44, September 13, 2005, New South Wales.

Hildreth, C. (1954): Point Estimates of Ordinates of Concave Functions, *Journal of the American Statistical Association*, 49, 598-619.

Hsieh, C. (2000): Measuring Biased Technological Change,.Princeton University, mimeo.

Ivaldi, M., N. Ladoux, H. Ossard and M. Simioni (1996): Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers. *Journal of applied econometrics*, 11(6), pp.649-667.

Japan Economic Newswire (1999): US Raps Japan on Car Fuel Efficiency Rules. Publication of March 8, 1999.

Joint Senate Resolution (2001): Joint Senate Resolution 2nd extension session. Bill number SJRX2-1, California. Introduced by Senator Karnette, filed with Secretary of State June 27, 2001 and adopted in Senate June 25, 2001.

Kamstra, M., Kramer, L and M. Levi (2000): Losing sleep at the market: The daylight saving anomaly. *American Economic Review*, 90(4).

Kemp, D. (2003): Australia Moves Closer to Kyoto Target. The Australian Government Department of the Environment and Heritage Media Release, 18 September, 2003.

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

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.

Koop, G., J. Osiewalski and M.F.J. Steel (1994): Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function. *Journal of Business and Economic Statistics*, 12(3) pp.339-46.

Koop, G., J. Osiewalski and M.F.J. Steel (1997): Bayesian Efficiency Analysis through Individual Effects: Hospital cost frontiers. *Journal of Econometrics*, 76, pp.77-105.

Kumbhakar, S.C. and E.G. Tsionas (*forthcoming*): Measuring Technical and Allocative Inefficiency in the Translog Cost System: A Bayesian Approach, *Journal of Econometrics*.

Kyodo News (2005): 140 Lawmakers to Submit Daylight Saving Time Bill. *Kyodo News* from March 17th, 2005.

Lambe, M. and P. Cummings (2000): The shift to and from daylight savings time and motor vehicle crashes. *Accident Analysis and Prevention* 32, 609-611.

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.

Lawson, L (2001): Testimony of Linda Lawson, Acting Deputy Assistant Secretary for Transportation Policy, U.S. Department of Transportation, before the House Science Committee, Energy Subcommittee, concerning daylight saving time and energy conservation, 24 May, 2001.

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

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.

Mas-Colell, A., M.D. Whinston and J.R. Green (1995): Microeconomic Theory. Oxford University Press.

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.

Meyer, B. (1995): Natural and Quasi-Experiments in Economics, *Journal of Business & Economic Statistics* 13, 151-161.

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

NEMMCO (2005): An Introduction to Australia’s National Electricity Market. National Electricity Market Management Ltd.

Newey, W. and K. West (1987): A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703-708.

O’Donnell, C.J., C.R. Shumway and V.E. Ball (1999): Input Demands and Inefficiency in U.S. Agriculture. *American Journal of Agricultural Economics*, 81, November, pp.865-880.

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

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

Outhred, H. (2006): Email communication with Hugh Outhred, Director, Center for Energy and Environmental Markets and Head, Energy Systems Research Group, School of Electrical Engineering & Telecommunications, Sydney, July 20, 2006.

Pepin, N. (1997): Time for a Change? *Area*, 29(1), 57-71.

Petersen, M.A. (2006): Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. National Bureau of Economics Research (NBER) Working Paper 11280.

Prerau, D. (2005): Seize the Daylight: The Curious and Contentious Story of Daylight Saving Time. Thunder's Mouth Press, New York.

Rock, B. (1997): Impact of daylight saving time on residential energy consumption and cost. Energy and Building, 25, 63-68.

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 Revisted: New Evidence From Micro Data. *Economic Studies*, 97-4.

Sayers, C. and Shields, D. 2001, Electricity Prices and Cost Factors, Productivity Commission Research Paper, AusInfo, Canberra.

Simon, C.P. and L. Blume (1994): Mathematics for Economists. W.W. Norton, New York.

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

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

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

The Japan Times (2004): Daylight-saving time wins support. *The Japan Times*, 26 November, 2004.

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

Turton, H. and Hamilton, C. (2001): Comprehensive emissions per capita for industrialised countries, The Australia Institute, September 2001.

US Hearing (2001a): Congressional Perspectives on Electricity Markets in California and the West and National Energy Policy. Hearing 107-8 before the subcommittee on energy and air quality of the Committee on Energy and Commerce House of Representatives 107th Congress, First Session, March 6, 2001.

US Hearing (2001b): Congressional Perspectives on Electricity Markets in California and the West and National Energy Policy. Hearing 107-30 before the subcommittee on energy and air quality of the Committee.

Vanduffel, S. (2005): Comonotonicity: From Risk Measurement to Risk Management. *Academisch Proefschrift*. Faculteit der Economische Wetenschappen en Econometrics, Amsterdam.

Varughese, J. and R.P. Allen (2001): Fatal accidents following changes in daylight savings time: the American experience. *Sleep Medicine* **2**, 31-36.

Wolff, H., T. Heckelei, and R.C. Mittelhammer (2004): Imposing Curvature and Monotonicity on Flexible Functional Forms: An Efficient Regional Approach. Working Paper.

Wooldridge, J.M. (2003): Cluster-Sample Methods in Applied Econometrics. *American Economic Review*, Vol. 93, 133-138.

Young, T. (2005): Independent Electricity System Operator, Ontario, Spokesperson, in Peter Gorrie: Get set for darker November mornings, *The Toronto Star*, 21 July, 2005, p. A1.

Zellner, A. (1962): An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* 57, 348–368.

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

Appendices

The appendix is divided into two subchapters. Appendix 1 has material that belongs to chapter 2 and Appendix 2 belongs to chapter 4.

Appendix 1

Appendix 1A: 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 $\mathbf{b}^{(*)} \in \mathbf{\Theta}$, the orthant of strictly positive prices π can always be partitioned into two disjoint subsets, $\boldsymbol{\pi}^{\mathrm{R}}|\mathbf{b}^{(*)} \cup \boldsymbol{\pi}^{\mathrm{IR}}|\mathbf{b}^{(*)} = \boldsymbol{\pi}$. 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:

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

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

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

$$\begin{array}{ll} p_1^{\mathrm{B}} \le p_1^* & \{\text{or } p_1^{\mathrm{B}} \ge p_1^*\} \\ p_2^{\mathrm{B}} \le p_2^* & \{\text{or } p_2^{\mathrm{B}} \ge p_2^*\} \\ : & \quad : \\ p_K^{\mathrm{B}} \le p_K^* & \{\text{or } p_K^{\mathrm{B}} \ge p_K^*\} \end{array}$$

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

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 property* I *and property* II *hold. Iff* $\mathbf{z} = [\, p_1^{\min\{\max\}}, p_2^{\min\{\max\}}, ..., p_K^{\min\{\max\}}\,]^{\mathsf{T}} \in \boldsymbol{\psi}_h^{\mathrm{R}}$, *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 $i_h(\mathbf{p}^{\mathrm{B}}) \le i_h(\mathbf{p}^*)$, hence $\mathbf{p}^{\mathrm{B}} \in \mathbf{B}^{\mathrm{IR}}$. From *property* II it follows that $\exists$ one vertex point $\mathbf{z} = [z_1, z_2, \ldots, z_K]^{\mathsf{T}}$ with the following property:

---

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

$z_1 \leq p_1^{\mathrm{B}}$ {or $z_1 \geq p_1^{\mathrm{B}}$}
$z_2 \leq p_2^{\mathrm{B}}$ {or $z_2 \geq p_2^{\mathrm{B}}$}
: :
$z_K \leq p_K^{\mathrm{B}}$ {or $z_K \geq p_K^{\mathrm{B}}$}

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

Since – under the conditions *property* I and *property* II – whenever $[\, p_1^{\min\{\max\}}, p_2^{\min\{\max\}}, ..., p_K^{\min\{\max\}}\,]^{\mathsf{T}} \in \boldsymbol{\psi}_h^{\mathrm{R}}$, then $\boldsymbol{\psi}_h^{\mathrm{R}} = \boldsymbol{\psi}$, we conclude that only this single vertex point has to be checked.[2] 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* $\partial i_h/\partial p_{\text{-m}}$ *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 × 1 vector $\mathbf{p}_{-m}^*$ and similarly partition $\mathbf{p}^{\mathrm{B}} \in \mathbf{B}$ into p_m^{B} and $\mathbf{p}_{-m}^{\mathrm{B}}$.

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

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

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

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

[2] 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.

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.

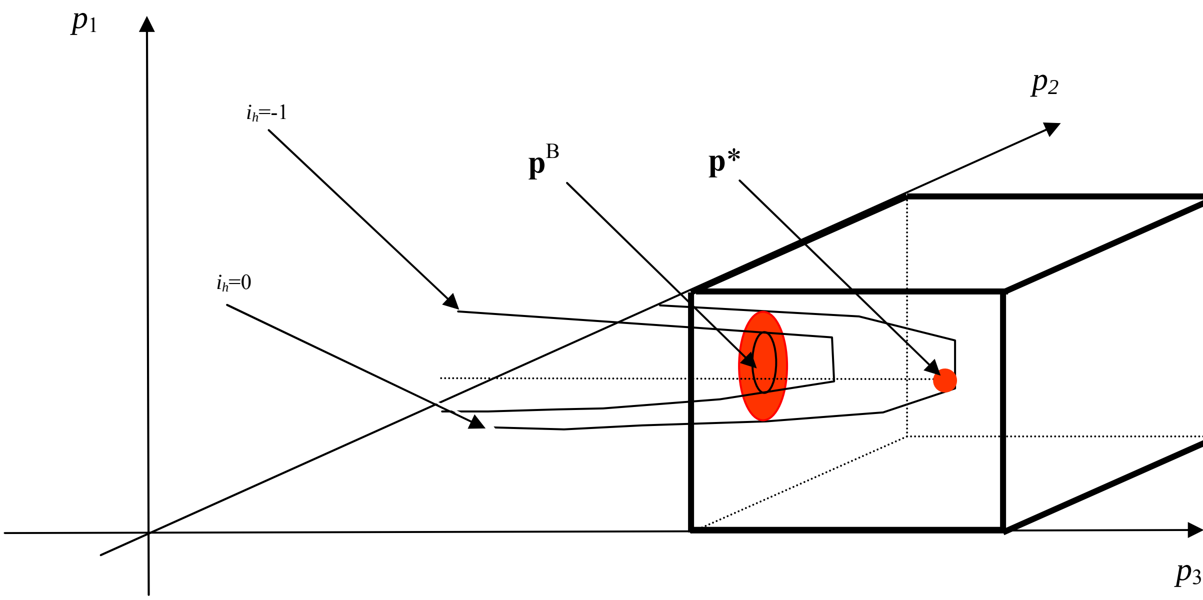
Corollary 2b**:** *Fix the* m^{th} *price axis from property III. Let* $\mathbf{S} \subset \mathbf{B} \subset \boldsymbol{\psi}$ *be that side of the hyperrectangle, which is orthogonal to the* m^{th} *price-axis and for which* p_m^{S} = $p_m^{\min\{\max\}}$ $\forall$ $(p_m^{\mathrm{S}}, \mathbf{p}_{-m}^{\mathrm{S}})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

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

In other words, if *property* II and III hold, then it is only necessary to evaluate **S** which is 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.

Figure A1: Inequality Constraint Function Level Sets i_h = -1 and i_h = 0 in price space π

If *property* II and *property* III hold, **p*** is irregular, and $\partial i_h/\partial p_3 \geq 0$, then the boundary side **S** facing towards the p_1–p_2 level contains irregular points $\mathbf{p}^B \in \mathbf{S}^{IR} \subset \mathbf{S}$.



$\mathbf{S}^{IR}$ is shaded in red. The set $\psi \subset \pi$ is indicated by the cube.

The following proposition provides sufficiency conditions to check only the extreme points $\mathbf{Z}_h^e$ of a convex set $\boldsymbol{\psi}$.[3] 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 $\mathbf{Z}_h^e$ is equivalent to the 2^K vertices defined in section 3.1 as $\mathbf{Z}_h$.[4]

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

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. $\boldsymbol{\psi}_h^R = \{\mathbf{p}: i_h \geq 0, \mathbf{p} \in \boldsymbol{\psi}\}$ is an

[3] $\mathbf{z}^e$ is an extreme point of $\boldsymbol{\psi}$ iff $\mathbf{z}^e = \lambda \cdot p_1 + (1-\lambda)p_2$, $\forall$ $p_1, p_2 \in \boldsymbol{\psi}$, $\lambda \in (0, 1)$, implies $\mathbf{z}^e = p_1 = p_2$.

[4] 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.

upper contour set $\mathbf{U}_0$ evaluated at $\omega = i_h = 0$ such that $\mathbf{Z}_h^e \in \boldsymbol{\psi}_h^R$ (by assumption). Since, by *property* IV, $\boldsymbol{\psi}$ is convex it follows that $\boldsymbol{\psi}_h^R = \mathbf{U}_0 \cap \boldsymbol{\psi}$ is convex (since the intersection of convex sets is convex). Finally, since any convex set is connected and $\mathbf{Z}_h^e \in \boldsymbol{\psi}_h^R$, it follows that $\boldsymbol{\psi}_h^R = \boldsymbol{\psi}$. ***Q.E.D.***

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

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

***Proof of Proposition 4*:** Suppose not, then $\exists$ $\mathbf{p}^* \in \boldsymbol{\psi}^{IR} \backslash \mathbf{S}^*$ for which either (a) 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 $\mathbf{p}^*$, $\mathbf{H}|_{\mathbf{p}^*}$, does not maintain the correct semi-definiteness. Again, by the property of homogenous functions we have $\partial f^2(t\mathbf{p}^*)/\partial\mathbf{p}\partial\mathbf{p}' = t^{\alpha-2}\partial^2 f(\mathbf{p}^*)/\partial\mathbf{p}\partial\mathbf{p}'\ \forall\ t > 0$. Since $\mathbf{H}|_{t\mathbf{p}^*}$ only differs from $\mathbf{H}|_{\mathbf{p}^*}$ by the multiple $t^{\alpha-2}$ the definiteness of the matrices is identical, hence $t\mathbf{p}^* \in \boldsymbol{\psi}^{\mathrm{IR}}\ \forall\ t > 0$. Consequently, the point $\mathbf{p}^{\mathbf{S}^*} \in \mathbf{S}^* \cap l(\mathbf{0},\mathbf{p}^*)$ is also irregular.

Q.E.D.

Appendix 1B: Proof of lemma 1 and proposition 5 to 6

Proof of Lemma 1: The proof follows immediately from the definition of $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}^* = \{\boldsymbol{\beta}: \mathbf{i}(\mathbf{p};\boldsymbol{\beta}) \geq \mathbf{0}\ \forall\ \mathbf{p} \in \boldsymbol{\psi}^*, \boldsymbol{\beta} \in \boldsymbol{\Theta}\}$ which implies that ceteris paribus, the larger the constraining set $\boldsymbol{\psi}^* \subset \boldsymbol{\pi}$, the smaller is the support $\boldsymbol{\Theta}^{\mathrm{R}}$, i.e. if $\boldsymbol{\psi}_1^* \subset \boldsymbol{\psi}_2^*$, then $\boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_1^* \supset \boldsymbol{\Theta}^{\mathrm{R}}|\boldsymbol{\psi}_2^*$. 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 proposition 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 $\boldsymbol{\Theta}^{\mathrm{R}}$.

Appendix 1C: Input price observations and out of sample points used for experiment II

Table 1C: 26 × 3 input price observation matrix P

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

$C =$ 4 scenario input price vectors

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

Appendix 2

Appendix 2A: Climate

Table A1: Historical Weather, Sunrise and Sunset data

Melbourne (VIC)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Celsius	20	20	19	16	13	11	10	11	13	15	17	19
Rainfall in mm	50	45	50	55	55	50	50	50	60	65	60	60
Average Sunrise	06:15	06:50	07:15	06:45	07:15	07:30	07:30	07:00	06:20	05:30	06:00	05:55
Average Sunset	20:45	20:20	19:40	17:50	17:20	17:05	17:20	17:45	18:10	18:40	20:10	20:40
Time: GMT+	11	11	11	10	10	10	10	10	10	10	11	11

Sydney (NSW)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Celsius	23	23	21	19	15	13	12	14	16	18	20	21
Rainfall in mm	100	110	130	120	120	125	100	75	65	75	80	75
Average Sunrise	06:00	06:30	06:55	06:20	06:40	07:00	07:00	06:30	05:50	07:15	05:40	05:40
Average Sunset	20:10	19:50	19:15	17:30	17:00	16:50	17:00	17:30	17:45	18:10	19:40	20:00
Time: GMT+	11	11	11	10	10	10	10	10	10	10	11	11

Adelaide (SA)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Celsius	20	20	19	17	15	12	12	12	14	16	17	21
Rainfall in mm	20	20	25	40	65	70	70	60	50	45	30	25
Average Sunrise	06:20	06:50	07:15	06:40	07:00	07:20	07:20	06:50	06:20	05:30	06:00	05:55
Average Sunset	20:30	20:10	19:35	17:50	17:20	17:10	17:20	17:45	18:05	18:30	20:00	20:25
Time: GMT+	10.5	10.5	10.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5	10.5	10.5

All sunrise/sunset hours are displayed in clock time (typical DST schedule), GMT: Greenwich Mean Time
Source: auinfo PTY LTD, Hornsby, NSW

Appendix 2B: Data Processing

Electricity data[5] are missing for occasional half-hours. We estimated the missing observations via interpolation using adjacent half hours. Hourly weather data are also missing for some occasional hours as well for four entire days (none of which fall in within 27 August – 29 October in any year, except for the air pressure variable). Hourly unobserved data were interpolated using adjacent hours. To estimate hourly weather in unobserved days, we applied a regression analysis which used information from the daily-level data set. Details and code for this procedure can be obtained from the authors upon request.

Schedules for most school vacations, state holidays, and federal holidays were obtained from the Australian Federal Department of Employment and Workplace Relations, The Department of Education and Children's Services (SA), and The Department of Education and Training (VIC). For years in which information was not available from the above institutions, the dates were obtained by internet search.

Federal holidays in Australia include Australia Day, Good Friday, New Years Day, Easter Monday, Boxing Day, Anzac Day, and the Queen’s Birthday. In years when Boxing Day and Anzac Day were moved to a different weekday than usual, both the original and the rescheduled holidays were modeled as holidays. State-specific holidays include Labor Day, the Melbourne Cup Day, and the Adelaide Cup Day. Public school vacations include Christmas break, Easter break, Winter break and Spring break.

[5] The NEMMCO data can be downloaded at http://www.nemmco.com.au/data/aggPD_2000to2005.htm.

Employment data are obtained from the Australian Bureau of Statistics, the Labor Force Spreadsheets, Table 12, using the series on the total number of employed persons by state for each quarter of the year.[6]

Sunrise, sunset, and twilight data were sourced from the U.S. Naval Observatory.[7] These data were then used to calculate the percentage of daylight and twilight in each half hour from January 1, 1999 to December 31, 2005 for Sydney, Melbourne, and Adelaide. Finally, we obtained the days and times of switches to and from DST from the Time and Date AS Company, located in Norway.[8]

While our data are provided in standard time, we conduct our analysis in nominal clock time. We therefore need to convert our data to clock time, which, for most affected observations, requires a simple one-hour shift. However, at the start of a DST period, the 02:00-03:00 interval (in clock time) is missing. To avoid a gap in our data, we duplicate the 01:30-02:00 information into the missing 02:00-02:30 half hour, and likewise equate the missing 02:30-03:00 period to our 03:00-03:30 observation. Further, when the DST period terminates, the 02:00-03:00 period (in clock time) is observed twice. Because our model is designed for only one observation in each hour, we average these dual observations.

Throughout the paper, several times we compare dates in Australia to equivalent dates in the northern hemisphere: In terms of sunrise sunset hours, the usual Australian DST starting date—the last Sunday in October—would

---

[6] For the employment data we used the series IDs A163206C, A163563A, A163257C, A163308T and A163359T.

[7] The astronomical data may be downloaded from http://aa.usno.navy.mil/.

[8] "Time and Date AS Company" provides data online at http://www.timeanddate.com/worldclock/timezone.html?n=240&syear=1990.

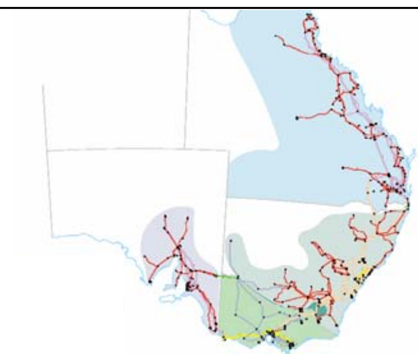
approximately correspond to the last Sunday in April on an equivalent latitude in the northern hemisphere. Equivalently, the date of the 2000 DST start in NSW and VIC (the last Sunday in August) corresponds *approximately* to the last Sunday in February in the northern hemisphere. Note, however, that the south latitude versus north latitude comparison can only be of an 'approximate' nature. Seasons are observed differently due to the fact that the earth is tilted toward the elliptic orbit in 23.5 degrees and the distance of the earth to the sun is not constant. This results into the following: on the dates of winter and summer solstices as well as the spring and fall equinoxes, the times of sunrise and sunset at a given latitude-longitude coordinate at the southern hemisphere are the same with the sunrise and sunset pattern at the same northern hemisphere latitude-longitude coordinate. However at all other dates, the sunrise-sunset times are slightly off, with differences increasing up to 15 minutes about 30 to 40 days after the equinox. Note that this approximation problem reduces with the dates of introducing DST earlier into the spring as the current DST switching dates discussed are closer to the equinox.[9]

[9] For example, 36 days after the spring equinox (i.e. corresponding to the usual start of DST in VIC around 28 October) Melbourne, at latitude 37.8 south and longitude 144.6 east observes sunrise and sunset at 19:17 and 08:52 UTC respectively. At the northern hemisphere, by contrast, 36 days after equinox (corresponding to about 27 April) sunrise-sunset at the corresponding latitude 37.8 north and longitude 144.6 east was at 19:31 and 09:06 UTC respectively. So while the total number of the daylight hours is the same, the time of daylight is shifted by around 14 minutes.

Appendix 2C: Information on Australia and the electricity market

Figure C1: Population density of Australia in the year 2004

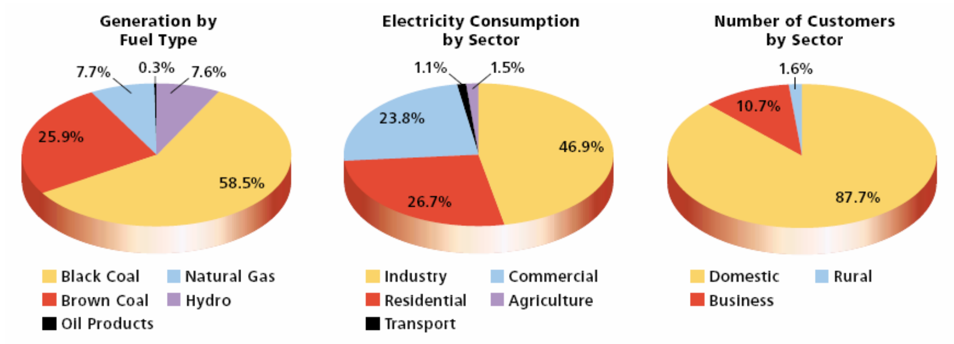
Figure C2: Electricity Grid



Source: NEMMCO, 2005

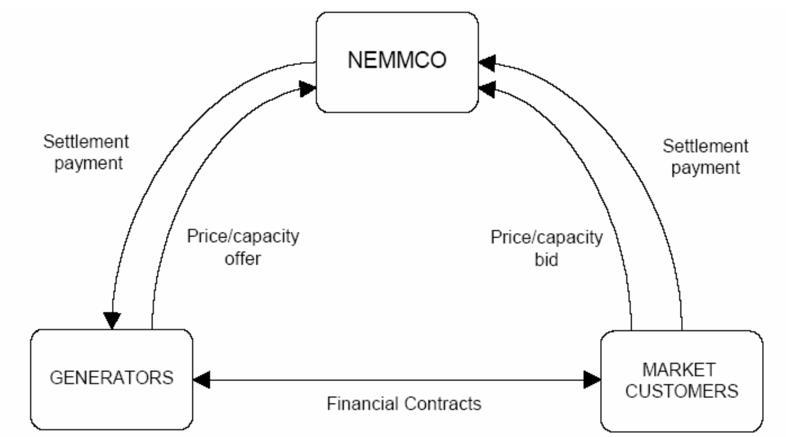
Figure C2 maps the world's longest interconnected power system, trading about 7 billion Australian dollars of electricity annually in the semi-privatized NEMMCO, serving about eight million end-use consumers. In this grid, 92% of the electricity produced relies on the burning of fossil fuels, and in total about 48% of the total per capita GHG emissions in Australia stem from the electricity sector (Kemp, 2003). Figure C3 displays the fuel mix in electricity production, and the split of consumption across economic sectors.

Figure C3: Electricity Production and Consumption in Australia



Source: NEMMCO, 2005

Figure C4: Settlement of Electricity Prices in the Electricity Market of VIC, NSW, QLD and SA



Source: Sayers, C. and Shields, 2001

Table C1: Characteristics of generators

Characteristic	Type			
	Gas and Coal-fired Boilers	Gas Turbine	Water (Hydro)	Renewable (Wind/Solar)
Time to fire-up generator from cold	8–48 hours	20 minutes	1 minute	dependent on prevailing weather
Degree of operator control over energy source	high	high	medium	low
Use of non-renewable resources	high	high	nil	nil
Production of greenhouse gases	high	medium-high	nil	nil
Other characteristics	medium-low operating cost	medium-high operating cost	low fuel cost with plentiful water supply; production severely affected by drought	suitable for remote and stand-alone applications; batteries may be used to store power

Source: NEMMCO, 2005

Appendix 2D: On Tourism to Australia

Figure D1 displays tourism data for VIC and SA, demonstrating that the 2000 Olympics did not significantly impact tourism in the third and fourth quarters of 2000. Tourism data for Sydney in NSW (Figure D2), however, shows that tourism increased in September 2000, and that there was no such increase in 1998 or 1999 (Australian Bureau of Statistics, 2001a, 2001b). Moreover, anecdotal evidence from Melbourne newspapers shows that Melbourne (the most frequently touristed location in VIC) did not experience any change in tourism before, during, or after the Olympic Games in 2000. Further details on tourism may be found in the Australian Bureau of Statistics' special report on Tourism related to the Olympics (2001b).

Figure D1: Quarterly Room Nights Occupied in VIC (left panel) and SA (right panel)

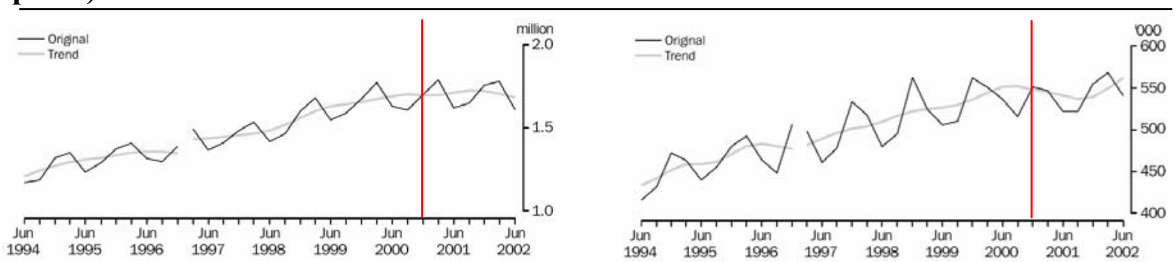
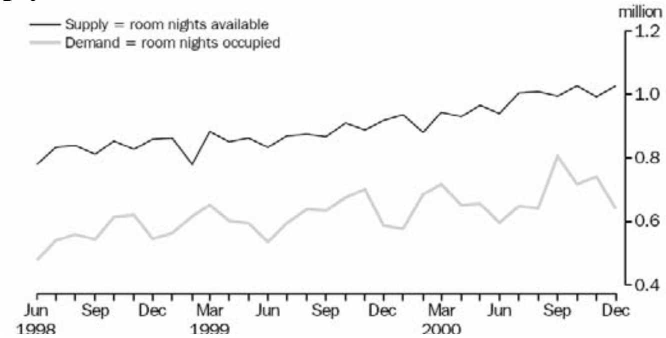


Figure: D2: Supply and Demand for Tourist Accommodations in Sydney



Source: Australian Bureau of Statistics, 2001. The vertical line indicates the 4th quarter in 2000 (December quarter). The treatment period "September" falls within the 3rd quarter 2000 and the treatment period "October" in the 4th quarter.

Appendix 2E: Estimation of Treatment Effect Model and Robustness

Table E1: Estimated treatment effects of the DST extension by half hour

Half hour beginning at	β_h	Std error	t-statistic	$\exp(\beta_h)$-1	Half hour beginning at	β_h	Std error	t-statistic	$\exp(\beta_h)$-1
00:00	-0.129	0.007	-18.24	-0.121	12:00	0.001	0.002	0.33	0.001
00:30	-0.012	0.007	-1.77	-0.012	12:30	0.000	0.002	0.19	0.000
01:00	0.019	0.007	2.75	0.019	13:00	-0.001	0.001	-0.71	-0.001
01:30	-0.050	0.006	-7.66	-0.048	13:30	-0.006	0.001	-4.72	-0.006
02:00	-0.045	0.007	-6.81	-0.044	14:00	-0.003	0.001	-2.48	-0.003
02:30	0.055	0.006	8.53	0.057	14:30	0.009	0.002	5.25	0.009
03:00	0.076	0.006	12.10	0.079	15:00	0.013	0.003	5.31	0.013
03:30	0.073	0.006	11.31	0.075	15:30	0.010	0.003	3.08	0.011
04:00	0.068	0.007	10.27	0.071	16:00	0.008	0.004	2.09	0.008
04:30	0.057	0.006	8.77	0.059	16:30	0.009	0.005	1.97	0.009
05:00	0.045	0.006	7.19	0.046	17:00	0.002	0.005	0.41	0.002
05:30	0.032	0.006	5.16	0.033	17:30	-0.014	0.006	-2.32	-0.014
06:00	0.025	0.006	4.18	0.025	18:00	-0.027	0.007	-3.63	-0.026
06:30	0.019	0.006	3.23	0.019	18:30	-0.048	0.007	-6.48	-0.047
07:00	0.015	0.006	2.58	0.015	19:00	-0.066	0.007	-8.84	-0.064
07:30	0.079	0.006	12.87	0.082	19:30	-0.055	0.008	-7.08	-0.054
08:00	0.077	0.006	12.70	0.080	20:00	-0.026	0.008	-3.33	-0.025
08:30	0.024	0.006	3.82	0.024	20:30	-0.008	0.008	-1.04	-0.008
09:00	0.006	0.005	1.23	0.006	21:00	-0.005	0.008	-0.62	-0.005
09:30	0.004	0.005	0.79	0.004	21:30	0.001	0.007	0.13	0.001
10:00	0.002	0.004	0.48	0.002	22:00	0.005	0.007	0.68	0.005
10:30	0.000	0.004	0.01	0.000	22:30	-0.006	0.007	-0.85	-0.006
11:00	0.003	0.003	1.06	0.003	23:00	-0.027	0.006	-4.33	-0.026
11:30	0.000	0.003	0.13	0.000	23:30	-0.124	0.007	-18.69	-0.117

Table E1 displays the estimated percentage impact of the DST extension on electricity demand in each half hour: these are the point estimates given by $\exp(\beta_h)$ - 1, and correspond to Figure 6. Note that the large effects in the late-night hours are caused by centralized off-peak water heaters in Melbourne (Outhred, 2006). These are triggered by timers set on Standard Time—groups of heaters are activated at 23:30 and 01:30. Each turns off on its own once its heating is complete. During the DST extension, each heater turns on one hour "late" (according to clock time). This drives the negative, then positive, overnight treatment effects.

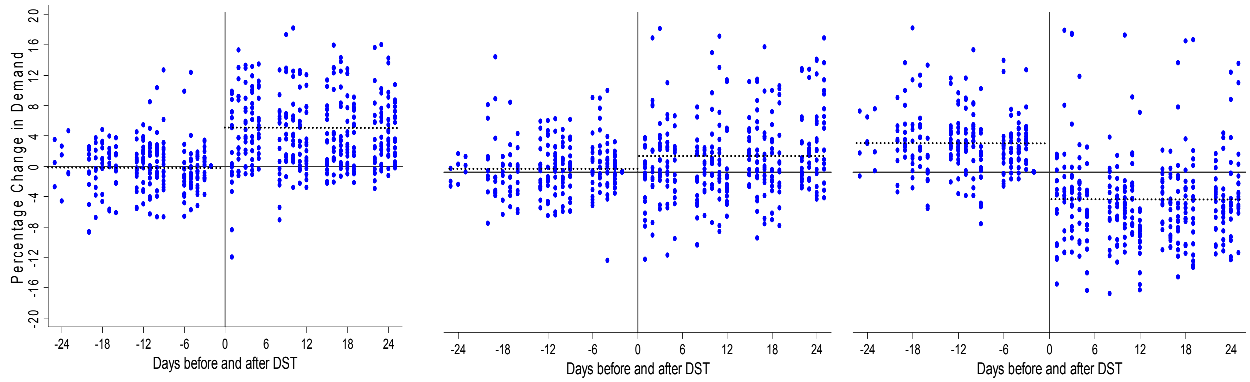
Justification of using 12:00 to 14:30 as the control period

Our estimation strategy uses the assumption that electricity demand in the afternoon is not affected by DST. The purpose of this subsection is to offer graphical and regression results to justify this assumption and to explain our specific choice of 12:00 to 14:30 as the base demand period for setting $\overline{q}$.

Figure E1 displays electricity demand for VIC and SA in 1999 and 2001-2005, one month before and one month after the late-October switch to DST in each year. Panel (a) indicates that morning demand increases immediately after the time change, while panel (c) shows that evening demand decreases. However, panel (b) demonstrates that afternoon demand is unaffected by the time change.

To verify the preliminary evidence offered by Figure E1, we perform a regression discontinuity analysis using the pre- and post-DST data in 1999 and 2001 to 2005, in both SA and VIC. The dependent variable is demand and the regressors consist of state and year fixed effects, their interaction, weather variables, a linear time trend, and a binary variable "DST" that is equal to one if DST is observed and zero otherwise.

Figure E1: Effect of DST on morning, afternoon and evening consumption



(a) morning 7:30-8:00 (b) afternoon 13:00-13:30 (c) evening 19:30-20:00

Vertical axis: Electricity demand relative to demand on the Friday preceding the start of DST. Each day contains a maximum of 12 data points (2 states over 6 years). Data excluded are: the year 2000, weekends, holidays, school holidays and "transition vacation days".

When we run this regression using only data from the morning hours of 7:30-8:00, we estimate that the coefficient on the DST variable is positive and significant: the point estimate is +121 with a standard error of 46. This agrees with the increase in morning demand shown in panel (a) of Figure E1. Similarly, we find that DST decreases evening demand: the point estimate during 19:30-20:00 is -103 with a standard error of 30.

During the afternoon, however, the estimated effect of DST is insignificant. Table E2 displays estimates of the DST coefficient, along with standard errors and *t*-values, for several afternoon half-hour intervals.[10] Our base period choice of 12:00–14:30 is driven by both the *t*-values shown and a desire to be conservative in our reference case estimate. While the lowest available *t*-value is for 13:00-13:30,

[10] Robustness checks for varying the sample size (changing the number of dates included before and after DST takes effect), using single hour equations or aggregating the hours did not yield results substantially different from those displayed in table E2.

suggesting that this would be an appropriate base period, its use yields a large estimate of the overall treatment effect θ: an increase in electricity consumption of 1.0%. To be more conservative in our final estimate, we instead report reference case results using 12:00-14:30 as the base period, even though the estimates reported in Table E2 suggest that DST may slightly increase electricity demand at this time. Despite this choice of base period, we still find a point estimate of θ that is positive, and reject prior studies' claims that extending DST conserves electricity.

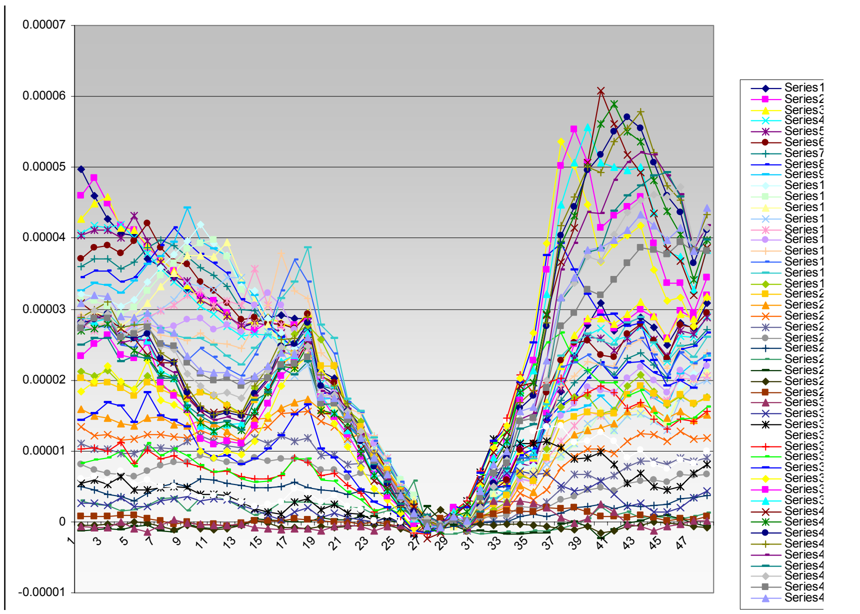
Table E2: Half-hourly DST effects on demand for VIC and SA

Halfhour	DST	std.error	t-value
11:00-11:30	40.19	45.89	0.88
11:30-12:00	34.22	46.43	0.74
12:00-12:30	42.05	46.11	0.91
12:30-13:00	36.33	47.14	0.77
13:00-13:30	13.28	48.74	0.27
13:30-14:00	19.41	51.08	0.38
14:00-14:30	46.83	51.70	0.91
14:30-15:00	59.03	52.00	1.14
15:00-15:30	53.46	52.77	1.01
15:30-16:00	43.28	52.08	0.83

The half hour from 13:00-13:30 exhibits the lowest t-value. The neighboring hours show monotonically increasing *t*-values respectively up to the period from 12:00-14:30 that is the base period used for $\overline{q}$.

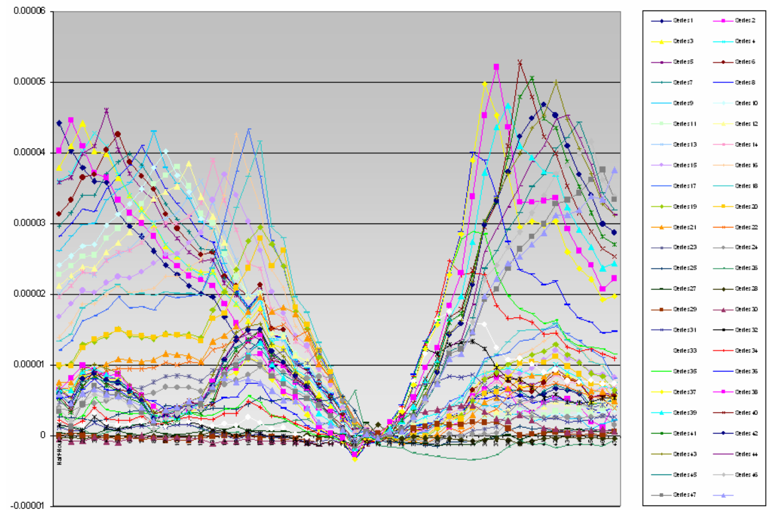
Figure E2 displays the covariance matrix of the treatment coefficients $\hat{\boldsymbol{\beta}}$ estimated from the reference case model. Each data series shown corresponds to the square root of the *h*th row of our estimated 48 x 48 clustered covariance matrix, **cov(**$\hat{\boldsymbol{\beta}}$**)**. The peak value of each series coincides with the diagonal-element var($\hat{\beta}_{hh}$). The off-diagonal elements become smaller with increasing distance from the diagonal element, because the dependency between neighboring half-hours decreases over time. The U-shaped pattern stems from the fact that the treatment effects between 12:00-14:30 have very small standard errors, by the design of the triple-DID method.

Figure E2: Illustration of the clustered covariance matrix of $\hat{\boldsymbol{\beta}}$



The estimated Newey-West covariance matrix is displayed in Figure E3. Here, the dependency between $\hat{\beta}_h$ and $\hat{\beta}_{h+i}$ declines more quickly than was the case with the clustered covariance because the Newey-West explicitly accounts for the serial correlation of $\boldsymbol{\varepsilon}$ so that the remaining covariance structure of $\hat{\boldsymbol{\beta}}$ exhibits less dependency among the neighboring half hours.

Figure E3: Covariance matrix estimated by Newey-West

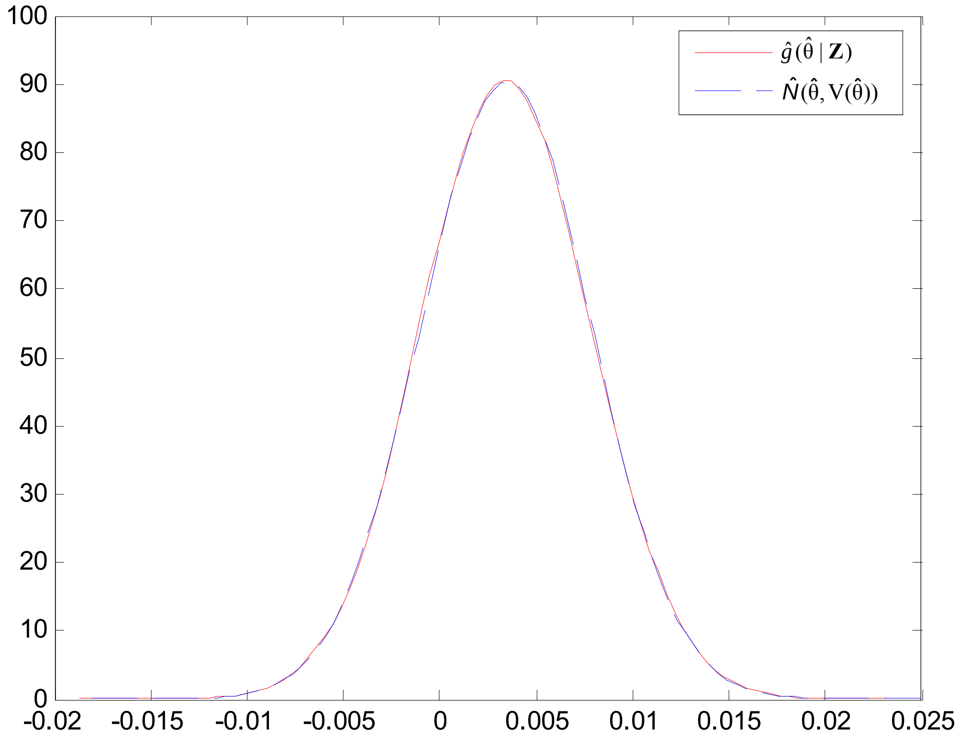


On the numerical equivalence between $\hat{g}(\hat{\theta} \mid \mathbf{Z})$ and $N(\hat{\theta}, \mathrm{V}(\hat{\theta}))$

In chapter 4.5 we approximate $g(\hat{\theta})$ by $N(\hat{\theta}, \mathrm{V}(\hat{\theta}))$. Figure E4 displays $\hat{g}(\hat{\theta} \mid \mathbf{Z})$ and $N(\hat{\theta}, \mathrm{V}(\hat{\theta}))$ in the case of the pooled treatment effect. Given the large sample, the close match between these two approaches justifies the approximation of the posterior $\hat{g}(\hat{\theta} \mid \mathbf{Z})$ with the simulated likelihood $\hat{g}(\hat{\theta})$ and the normal approximation $N(\hat{\theta}, \mathrm{V}(\hat{\theta}))$.[11]

[11] The equivalence of these results is driven by central limit theorem: the sum of the 48 non-iid lognormals is large enough relative to the dependency, so that the asymptotics take over.

Figure E4: Estimated density function $\hat{g}(\hat{\theta} \mid \mathbf{Z})$ and simulated normal density



	$\hat{g}(\hat{\theta} \mid \mathbf{Z})$ approximated by 100,000 draws	$N(\hat{\theta}, \mathrm{V}(\hat{\theta}))$ approximated by 100,000 draws	Analytical Estimates based on (3) and (4)
Mean	0.00343	0.00345	0.00341
Std	0.00433	0.00434	0.00432
Skewness	0.02274	0.00110	
Kurtosis	3.00578	3.00422	