Imposing Theoretical Regularity on Flexible Functional Forms*

Apostolos Serletis†
Department of Economics
University of Calgary
Calgary, Alberta, T2N 1N4

and

Guohua Feng
Department of Economics
University of North Texas
Denton, TX 76203

Forthcoming in: *Econometric Reviews*

April 18, 2013

*Serletis gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada (SSHRCC). Guohua Feng gratefully acknowledges financial support from the Australian Research Council.

†Corresponding author. Phone: (403) 220-4092; Fax: (403) 282-5262; E-mail: Serletis@ucalgary.ca; Web: http://econ.ucalgary.ca/serletis.htm

"Although unconstrained specifications of technology are more likely to produce violations of curvature than monotonicity, I believe that induced violations of monotonicity become common, when curvature alone is imposed. Hence the now common practice of equating regularity with curvature is not justified."

William A. Barnett (2002, p. 199)

Abstract:

In this paper we build on work by Gallant and Golub (1984), Diewert and Wales (1987), and Barnett (2002) and provide a comparison among three different methods of imposing theoretical regularity on flexible functional forms — reparameterization using Cholesky factorization, constrained optimization, and Bayesian methodology. We apply the methodology to a translog cost and share equation system and make a distinction between local, regional, pointwise, and global regularity. We find that the imposition of curvature at a single point does not always assure regularity. We also find that the imposition of global concavity (at all possible, positive input prices), irrespective of the method used, exaggerates the elasticity estimates and rules out the possibility of a complementarity relationship among the inputs. Finally, we find that constrained optimization and the Bayesian methodology with regional (over a neighborhood of data points in the sample) or pointwise (at every data point in the sample) concavity imposed can guarantee inference consistent with neoclassical microeconomic theory, without compromising much of the flexibility of the functional form.

JEL classification: C51; C3; C13.

Keywords: Flexible functional forms; Translog cost function; Monotonicity; Curvature.

1 Introduction

The widespread use of flexible functional forms in factor and consumer demand analysis has given researchers the ability to model technology and consumer preferences with no restrictions on the nature of the substitutability/complementarity relationship between pairs of goods. Unfortunately, however, theoretical regularity restrictions (of positivity, monotonicity, and curvature) automatically met by simpler forms, such as the Cobb-Douglas and the constant elasticity of substitution (CES), or by globally regular flexible functional forms, might not be satisfied with most flexible functional forms, and thus the researcher who desires to approximate arbitrary technology or preferences often ends up with an approximation that violates theoretical regularity.

The usefulness of flexible functional forms depends on whether they satisfy the theoretical regularity conditions of positivity, monotonicity, and curvature, and in the literature there has been a tendency to ignore theoretical regularity. According, for example, to Serletis and Shahmoradi (2007) only three out of fourteen studies in the monetary asset demand literature since 1983 have addressed theoretical regularity issues. Moreover, there is a tendency to remove 'regularity' from the empirical literature and replace it with 'curvature.' In fact, as Barnett (2002, p. 199) put it in his *Journal of Econometrics* Fellow's opinion article,

> "monotonicity is rarely even mentioned in that literature. But without satisfaction of both curvature and monotonicity, the second-order conditions for optimizing behavior fail, and duality theory fails. The resulting first-order conditions, demand functions, and supply functions become invalid."

In fact, in addition to the imposition of curvature, the imposition of monotonicity (when violated) is an important issue. Moreover, the imposition of curvature may induce violations of monotonicity which may not occur otherwise — see Barnett (2002, p. 202). As Barnett and Pasupathy (2003, p. 151) put it,

> "research on models permitting imposition of both curvature and monotonicity remains at an early stage and has so far had little impact on the literature on production modeling. While a difficult literature, we believe that research models permitting flexible imposition of true regularity — i.e., both monotonicity and curvature — should expand."

In this paper, we take up Barnett on his suggestion and move the literature forward, by providing a comparison among three different methods of imposing theoretical regularity — reparameterization using Cholesky factorization, constrained optimization, and Bayesian methodology. The reparameterization technique to the imposition of curvature is due to Wiley *et al.* (1973) and was pioneered in the flexible functional forms literature by Diewert and

Wales (1987) to impose global concavity on the generalized McFadden (also called normalized quadratic) cost function, and was later generalized by Ryan and Wales (2000) to impose local concavity (at a single point in the sample) on the translog and generalized Leontief cost functions. The second approach is nonlinear constrained optimization, pioneered by Gallant and Golub (1984) to impose concavity restrictions on the Fourier cost function, and recently used by Serletis and Shahmoradi (2005) to impose curvature on the Asymptotically Ideal Model (AIM) and Fourier globally flexible functional forms in the context of modelling the demand for monetary assets, and by Feng and Serletis (2008) in the context of the AIM cost function with technological change. The third approach of imposing theoretical regularity is the Bayesian approach, pioneered by Terrell (1996) and Griffiths *et al.* (2000), in imposing regularity on factor demand systems, and by Koop *et al.* (1997) in imposing regularity on stochastic frontier models.

We apply the methodology to a translog cost and share equation system, using annual KLEM (capital, labor, energy, and intermediate materials) data for total manufacturing in the United States, over the period from 1953 to 2001. In doing so, we make a distinction between 'local regularity' (at some representative point in the data), 'regional regularity' (over a neighborhood of data points in the sample), 'pointwise regularity' (at every data point in the sample), and 'global regularity,' occurring at all possible, positive input prices.

We use a cost function, because cost functions lie at the heart of basic economic theory and are more commonly used than other functions (i.e. production functions, profit functions, or distance functions), whether in the traditional factor demand literature or the more recent firm-efficiency related literature. We use the translog flexible functional form, because it is one of the most commonly used functional forms in the literature. Of course, our discussion applies to many other flexible functional forms such as, for example, the generalized Leontief and normalized quadratic, as well as to the literature of consumer preferences modelling if we replace the cost function with an indirect utility function.

The rest of the paper is organized as follows. In Section 2, we use duality theory to represent the production technology in terms of the cost function, present the (locally flexible) translog cost function, and specify the theoretical regularity conditions (of positivity, linear homogeneity, monotonicity, and curvature) required by neoclassical microeconomic theory. In Section 3 we present a stochastic specification of the model and in Section 4 we discuss three different methods for imposing the theoretical regularity conditions — reparameterization using Cholesky factorization, constrained optimization, and Bayesian methodology. In Section 5 we apply our methodology to annual KLEM (capital, labor, energy, and intermediate materials) data for total manufacturing in the United States and report and discuss the results. The last section summarizes and concludes the paper.

2 The Theoretical Framework

We consider a firm whose long-run production technology is given by

$$y = f(\boldsymbol{x}, t), \tag{1}$$

where y is output, f is a continuous twice differentiable nondecreasing and quasiconcave function of a vector of inputs $\boldsymbol{x} \geq \mathbf{0}$, and t denotes a technology index. The production technology can also be described in the dual setup (cost function), under certain conditions. In particular, if firms competitively minimize the cost of production subject to producing a given amount of output, then the technology (1) is completely described by the dual cost function

$$C = C(\boldsymbol{p}, y, t) = \phi(y) c(\boldsymbol{p}, t), \tag{2}$$

where C is total cost, $\phi(y)$ is a function of output, $\boldsymbol{p} > \mathbf{0}$ is the input price vector (with $\mathbf{0}$ being the null vector), and $c(\boldsymbol{p}, t)$ is the corresponding unit cost function — see Sickles (1985). As required by duality theory, the cost function is nondecreasing (i.e. monotonic), linearly homogeneous, and concave function of prices — see Diewert (1982) for an excellent review of duality theory.

2.1 The Translog Cost Function

The cost function in (2) can be approximated by a flexible functional form, such as, for example, the translog [see Christensen *et al.* (1975)], the generalized Leontief [see Diewert (1971)], the normalized quadratic [see Diewert and Wales (1987)], the Fourier [see Gallant (1982)], or the Asymptotically Ideal Model [see Barnett *et al.* (1991)]. In this paper, we use the translog flexible functional form, due to its widespread use in both the traditional factor demand systems literature and more recent firm efficiency literature.

One of the features of the KLEMS data set used in this paper is that constant returns to scale has been built in by the U.S. Bureau of Labor Statistics (for more details, see the description of the data in Section 5). To be consistent with this feature, we set $\phi(y) = y$. We further assume that the unit cost function $c(\boldsymbol{p}, t)$ takes on a translog functional form to obtain the following cost function

$$\ln C(\boldsymbol{p}, y, t) = \ln y + \beta_0 + \sum_{i=1}^{M} \beta_i \ln p_i + \beta_t t$$

$$+ \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^{M} \beta_{it} t \ln p_i + \frac{1}{2} \beta_{tt} t^2, \tag{3}$$

where M is the number of inputs. Symmetry requires $\beta_{ij} = \beta_{ji}$ $(i,\ j = 1, \cdots, M)$. There are two reasons for assuming constant returns to scale. The first is for simplicity, and the second is that constant returns to scale is already imposed in the data that we use in the empirical application in Section 5.

The factor share equations are obtained using Shephard's lemma and take the form

$$s_i = \frac{p_i x_i}{C} = \beta_i + \sum_{j=1}^{M} \beta_{ij} \ln p_j + \beta_{it} t, \quad i = 1, \cdots, M, \tag{4}$$

where s_i is the cost share for input i. It is to be noted that we assume that factor shares are affected by technical change; that is, the form of technical change that we assume here is not Hicks neutral.

2.2 Theoretical Regularity

As required by neoclassical microeconomics theory, the translog cost function has to satisfy the regularity conditions of positivity, homogeneity, monotonicity, and curvature. Positivity requires that the estimated cost be positive for all the data observations. Homogeneity requires that the translog cost function be linearly homogeneous in input prices. Monotonicity requires that the first-order derivatives of the cost function, $\partial C(\boldsymbol{p}, y, t)/\partial p_i$ $(i = 1, \cdots, M)$, which correspond to input demands, be nonnegative. Curvature requires that the cost function, $C(\boldsymbol{p}, y, t)$, be a concave function of prices or, equivalently, that the Hessian matrix of the cost function, $\boldsymbol{H}$, be negative semidefinite.

More formally, positivity will be satisfied if the estimated cost is positive,

$$\hat{C}(\boldsymbol{p}, y, t) > 0,$$

and linear homogeneity is satisfied if

$$\sum_{i=1}^{M} \beta_i = 1,\ \sum_{i=1}^{M} \beta_{ij} = \sum_{j=1}^{M} \beta_{ji} = \sum_{i=1}^{M} \beta_{it} = 0. \tag{5}$$

Monotonicity is satisfied if the estimated input shares are positive,

$$\hat{s}_i > 0, \quad i = 1, \cdots, M, \tag{6}$$

since both p_i $(i = 1, \cdots, M)$ and C are positive.

Finally, curvature is satisfied if the Hessian matrix

$$\boldsymbol{H} = \left[\frac{\partial^2 C(\boldsymbol{p}, y, t)}{\partial p_i \partial p_j}\right],$$

for i, $j = 1, \cdots, M$, is negative semidefinite. In fact, Diewert and Wales (1987) show that $\boldsymbol{H}$ is negative semidefinite if and only if the following matrix is negative semidefinite

$$\boldsymbol{G} = \boldsymbol{B} - \underline{\boldsymbol{s}} + \boldsymbol{s}\boldsymbol{s}', \tag{7}$$

where $\boldsymbol{B} = [\beta_{ij}]$, $\boldsymbol{s} = (s_1, \cdots, s_M)$ is the share vector, and $\underline{\boldsymbol{s}}$ is the $M \times M$ diagonal matrix which has the share vector $\boldsymbol{s}$ on the main diagonal.

It should be noted here that regularity should not be treated as being equivalent to curvature alone; instead it includes all the three above conditions, namely, positivity, monotonicity, and curvature. See Barnett (2002). For the purpose of conceptual clarity and notational simplicity, we make a distinction between local, regional, pointwise, and global regularity imposition, according to the way regularity is imposed:

Definition 1 *Local regularity imposition. It involves imposing regularity at a single reference point in the sample.*

Definition 2 *Regional regularity imposition. It involves imposing regularity over a 'neighborhood' of data points in the sample.*

Definition 3 *Pointwise regularity imposition. It involves imposing regularity at all data points in the sample.*

Definition 4 *Global regularity imposition. It involves imposing regularity in such a way that regularity is satisfied at all possible, positive input prices.*

Let's consider curvature as an example to illustrate the above definitions. Local curvature imposition requires imposing negative semidefiniteness of $\boldsymbol{G}$ at a single chosen reference point in the sample. Although it guarantees concavity at one point only, it may well be that a judicious choice of reference point leads to satisfaction of concavity at other (or all) data points in the sample. Regional curvature imposition requires imposing negative semidefiniteness of $\boldsymbol{G}$ over a chosen neighborhood of data points in the sample. Like with the case of local curvature imposition, a judicious choice of a neighborhood of data points may lead to satisfaction of concavity at all data points in the sample. Pointwise curvature imposition requires imposing negative semidefiniteness of $\boldsymbol{G}$ at all data points in the sample. Global curvature imposition requires $\boldsymbol{B}$ to be negative semidefinite — see Diewert and Wales (1987). The fact that the sample is stochastically generated by the economy suggests that a regularity requirement at a single data point or over a neighborhood of data points is not very appealing, unless it ensures regularity at every data point in the sample.

3 Empirical Modelling

In order to estimate equation systems such as (3) and (4), a stochastic component, $\boldsymbol{\epsilon}$, is added to the equations as follows

$$\boldsymbol{q}_t = \boldsymbol{v}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \qquad t = 1, ..., T \tag{8}$$

where $\boldsymbol{q}_t = (q_{1t}, \cdots, q_{Nt})'$, with $N = M + 1$, $q_{it} = s_{it}$ $(i = 1, \cdots, N - 1)$ and q_{Nt} is equal to $\ln C(\boldsymbol{p}, y, t)$ measured at time t. $\boldsymbol{v}_t$ is given by the right-hand side of the translog cost and share equations system, equations (4) and (3). $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_N)'$, with $\boldsymbol{\beta}_i$ $(i = 1, \cdots, N)$ being the corresponding vector of parameters for v_i $(i = 1, \cdots, N)$. Finally, $\boldsymbol{\epsilon}_t$ is a vector of stochastic errors and we assume that $\boldsymbol{\epsilon}_t \sim$Normal$\left(\mathbf{0}, \widetilde{\boldsymbol{\Omega}}\right)$, where $\mathbf{0}$ is a null vector, $\widetilde{\boldsymbol{\Omega}}$ is the $N \times N$ error covariance matrix.

It should also be noted that the shares sum to unity and the random disturbances corresponding to the M share equations thus sum to zero. This yields a singular covariance matrix of errors. Barten (1969) has shown that full information maximum likelihood estimates of the parameters can be obtained by arbitrarily deleting any one equation; the resulting estimates are invariant with respect to the equation deleted. In the empirical part of this paper, we arbitrarily delete the last (i.e. the N^{th}) share equation.

4 Methods of Imposing Theoretical Regularity

As mentioned above, positivity is usually automatically satisfied and thus is not the focus of this paper. In what follows we discuss three methods of imposing the theoretical regularity conditions of monotonicity and curvature, which are more commonly found to be violated.

4.1 Cholesky Factorization

First used by Wiley *et al.* (1973) and then popularized by Diewert and Wales (1987), the Cholesky factorization approach is mainly used to impose curvature when it is not satisfied. For the case of concavity, this can be done by expressing the Hessian matrix or a transformed matrix of the Hessian matrix (e.g., $\boldsymbol{G}$ in our particular case) as the negative of the product of a lower triangular matrix and its conjugate transpose. As shown below, there are two things that are worth noting. First, while the Cholesky factorization approach is only used to impose curvature in the previous literature, it can also be used to impose positivity and monotonicity. Second, for the case of curvature, this approach is capable of imposing local (at a single data point) and global (at all possible, positive input prices) curvature; however, it is incapable of imposing regional (over a neighborhood of data points in the sample) or pointwise (at every data point in the sample) curvature. Third, for the cases of monotonicity and positivity, this approach is only capable of imposing local monotonicity and positivity conditions.

We first discuss curvature. For the case of local curvature, we set all prices and output to unity and normalize t to zero at the reference point where concavity will be imposed. Then, at the reference point, the matrix $\boldsymbol{G}$ in (7) can be simplified as follows

$$\boldsymbol{G}_{ij} = \beta_{ij} + \beta_i \beta_j - \delta_{ij}\beta_i, \tag{9}$$

with $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. As noted above, concavity at the reference point requires $\boldsymbol{G}$ be negative semidefinite.

As in Feng and Serletis (2008) and Ryan and Wales (2000), we exploit the key idea behind the Cholesky factorization approach and set $\boldsymbol{G} = -\boldsymbol{K}\boldsymbol{K}'$ as follows

$$\beta_{ij} + \beta_i \beta_j - \delta_{ij}\beta_i = \left(-\boldsymbol{K}\boldsymbol{K}'\right)_{ij} \quad i, j = 1, \cdots, M, \tag{10}$$

where $\boldsymbol{K}$ is a lower triangular matrix. Noting that $\sum_{j=1}^{M} \beta_{ij} = 0$ and $\sum_{j=1}^{M} \beta_j = 0$ (see (5)), it can be easily shown that

$$\sum_{j=1}^{M} \boldsymbol{G}_{ij} = \sum_{j=1}^{M} (\beta_{ij} - \beta_i \delta_{ij} + \beta_i \beta_j) = 0, \tag{11}$$

i.e. the elements in the same row of $\boldsymbol{G}$ add to zero. Further, (11) implies the following restriction on the elements of $\boldsymbol{K}$

$$\sum_{i=1}^{M} k_{ij} = 0, \qquad j = 1, \cdots, M, \tag{12}$$

i.e., the elements in the same column of $\boldsymbol{K}$ add to zero. (12) can be easily shown by expanding out $\boldsymbol{G} = -\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{G}$ satisfies (11). Combining (10) and (12), we can replace the elements of $\boldsymbol{B} = [\beta_{ij}]$ by those of $\boldsymbol{K}$.

For the case with three inputs ($M = 3$), equations (10) and (12) imply the following restrictions on the elements of $\boldsymbol{K}$

$$\beta_{11} = -k_{11}^2 + \beta_1 - \beta_1^2 = -(k_{21} + k_{31})^2 + \beta_1 - \beta_1^2; \tag{13}$$
$$\beta_{12} = -k_{11}k_{21} - \beta_1\beta_2 = (k_{21} + k_{31})k_{21} - \beta_1\beta_2; \tag{14}$$
$$\beta_{13} = -k_{11}k_{31} - \beta_1\beta_3 = (k_{21} + k_{31})k_{31} - \beta_1\beta_3; \tag{15}$$
$$\beta_{22} = -(k_{21}^2 + k_{22}^2) + \beta_2 - \beta_2^2 = -k_{21}^2 - {k_{32}}^2 + \beta_2 - \beta_2^2; \tag{16}$$
$$\beta_{23} = -(k_{21}k_{31} + k_{22}k_{32}) - \beta_2\beta_3 = -k_{21}k_{31} + k_{32}^2 - \beta_2\beta_3; \tag{17}$$
$$\beta_{33} = -(k_{31}^2 + k_{32}^2 + k_{33}^2) + \beta_3 - \beta_3^2 = -(k_{31}^2 + k_{32}^2) + \beta_3 - \beta_3^2,$$

which guarantee concavity of the cost function at the reference point and may also induce concavity of the cost function at other data points.

For the case of global curvature, the translog will satisfy concavity in prices globally if (assuming positive shares) the matrix $\boldsymbol{B} = [\beta_{ij}]$ is negative semidefinite — see Diewert and Wales (1987) . To impose this condition, we can again exploit the key idea behind the Cholesky factorization approach and in this case set $\boldsymbol{B} = -\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{K}$ is defined as above. In the case, for example, of three inputs, the following restrictions concerning the relationship between the elements of $\boldsymbol{B}$ and those of $\boldsymbol{K}$ can be obtained

$$\beta_{11} = -k_{11}^2 = -(k_{21} + k_{31})^2; \tag{18}$$
$$\beta_{12} = -k_{11}k_{21} = (k_{21} + k_{31})k_{21}; \tag{19}$$
$$\beta_{13} = -k_{11}k_{31} = (k_{21} + k_{31})k_{31}; \tag{20}$$
$$\beta_{22} = -(k_{21}^2 + k_{22}^2) = -k_{21}^2 - k_{32}^2; \tag{21}$$
$$\beta_{23} = -(k_{21}k_{31} + k_{22}k_{32}) = -k_{21}k_{31} + k_{32}^2; \tag{22}$$
$$\beta_{33} = -(k_{31}^2 + k_{32}^2 + k_{33}^2) = -(k_{31}^2 + k_{32}^2). \tag{23}$$

However, as shown in Lau (1978) and Diewert and Wales (1987), the imposition of global regularity will reduce the translog model to the Cobb-Douglas, which is not a flexible functional form. This has actually been confirmed by the predominance of zeros for the β_{ij} in the empirical part (see Section 5.2).

In addition to imposing local and global curvature, the Cholesky factorization approach can also be used for the imposition of local monotonicity, although this has not been done previously. In particular, at the same reference point where local curvature is imposed, the factor share equations in (4) can be rewritten as

$$s_i = \beta_i, \quad i = 1, \cdots, M, \tag{24}$$

since all prices are set to unity and the time trend is normalized to zero at this point when local curvature is imposed. To impose monotonicity locally (i.e. $s_i > 0$), we can again exploit the key idea behind the Cholesky factorization approach and in this case set $\beta_i = \gamma_i^2$ (a 1×1 version of the Cholesky decomposition). Equation (8) can then be estimated after replacing β_i with γ_i^2 and the elements of $\boldsymbol{B}$ with those of $\boldsymbol{K}$.

In a similar manner, it can also be used to impose local positivity, although this has not been done previously too. This is because at the reference point where local curvature is imposed, the cost equation can be expressed as

$$\ln C\left(\boldsymbol{p}, y, t\right) = \beta_0$$

since all prices are set to unity and the time trend is normalized to zero at this point when local curvature is imposed. Thus as with the case of monotonicity imposition, we can then set $\beta_0 = \widetilde{\gamma}_i^2$. Equation (8) can then be estimated after replacing β_0 with $\widetilde{\gamma}_i^2$, β_i with γ_i^2 and the elements of $\boldsymbol{B}$ with those of $\boldsymbol{K}$.

As noted above, while the Cholesky factorization approach is capable of imposing local and global curvature, it is incapable of imposing regional or pointwise curvature. This is because the imposition of regional or pointwise curvature involves the manipulation of both the parameter space (i.e., β_{ij} and β_i) and the data space (i.e., p_i, y, and t). However, as we have seen above [i.e., (10)–(17)], the Cholesky factorization approach is only capable of the manipulation of the parameter space (i.e. by setting all prices and output to unity and normalizing t to zero at the reference point). For the same reason, while the Cholesky factorization approach is capable of imposing local positivity and monotonicity, it is incapable of the imposition of regional, pointwise or global monotonicity, in that the sufficient condition for regional, pointwise or global monotonicity and positivity, which involves only the parameter space, does not exist for the translog functional form.

4.2 Nonlinear Constrained Optimization

As noted in the introduction, the nonlinear constrained optimization approach was introduced into the flexible functional forms literature by Gallant and Golub (1984) and has been used recently by Serletis and Shahmoradi (2005) and Feng and Serletis (2008). The TOMLAB/NPSOL tool box with MATLAB has been used in all the aforementioned papers, and is also used in this paper for monotonicity and curvature constrained optimization. TOMLAB/NPSOL uses a sequential quadratic programming algorithm and is suitable for both unconstrained and constrained optimization of smooth (that is, at least twice-continuously differentiable) nonlinear functions.

To reduce the complexity and computational time of our optimization problem and in order to simplify the objective function and constraints, we impose homogeneity by normalizing the cost and input prices by one of the input prices. This method of imposing homogeneity is widely used in this literature. In particular, we normalize input prices and the cost, C, by the last input price p_M, thereby transforming the cost and share equations system, (3) and (4), as follows

$$\ln \frac{C\left(\boldsymbol{p}, y, t\right)}{p_M} = \ln y + \beta_0 + \sum_{i=1}^{M-1} \beta_i \ln \frac{p_i}{p_M} + \beta_t t$$

$$+ \frac{1}{2} \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} \beta_{ij} \ln \frac{p_i}{p_M} \ln \frac{p_j}{p_M} + \sum_{i=1}^{M-1} \beta_{it} t \ln \frac{p_i}{p_M} + \frac{1}{2} \beta_{tt} t^2, \tag{25}$$

and

$$s_i = \beta_i + \sum_{j=1}^{M-1} \beta_{ij} \ln \frac{p_j}{p_M} + \beta_{it} t, \quad i,\ j = 1, \cdots, M-1. \tag{26}$$

Once the parameters in (25) and (26) are estimated, the parameters $\beta_M, \beta_{M1}, \beta_{M2}, \cdots, \beta_{MM}$, and β_{Mt} can be recovered by using the homogeneity conditions in (5).

In this case, the empirical model can be written as

$$\boldsymbol{w}_t = \boldsymbol{X}_t \widetilde{\boldsymbol{\beta}} + \boldsymbol{e}_t, \qquad t = 1, ..., T \tag{27}$$

where $\boldsymbol{w}_t = (w_{1t}, \cdots, w_{Mt})'$, with $w_{it} = s_{it}$ $(i = 1, \cdots, M-1)$ and w_{Mt} is equal to $\ln [C(\boldsymbol{p}, y, t) / p_M]$ measured at time t. $\boldsymbol{X}_i$ $(i = 1, \cdots, M-1)$ is given by the right-hand side of each of (26) and $\boldsymbol{X}_M$ is given by the right-hand side of (25). $\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\boldsymbol{\beta}}_1, \cdots, \widetilde{\boldsymbol{\beta}}_M\right)'$, with $\widetilde{\boldsymbol{\beta}}_i$ being the corresponding vector of parameters for $\boldsymbol{X}_i$. $\boldsymbol{e}_t$ is a vector of stochastic errors, which is assumed to follow the following distribution: $\boldsymbol{\epsilon}_t \sim$Normal$(\mathbf{0}, \boldsymbol{\Omega})$, where $\mathbf{0}$ is a null vector, $\boldsymbol{\Omega}$ is the $M \times M$ error covariance matrix.

To estimate (27) we specify the objective function, which is the log likelihood function in this paper. In particular, under the assumption that $\boldsymbol{e}_t$ is iid normal, the log likelihood function for a sample of T joint observations can be shown to be

$$\ln L\left(\boldsymbol{w} \middle| \widetilde{\boldsymbol{\beta}}, \boldsymbol{\Omega}\right) = -\frac{MT}{2} \log(2\pi) - \frac{T}{2} \log(\boldsymbol{\Omega}) - \frac{1}{2} \sum_{i=1}^{T} \boldsymbol{e}_t' (\boldsymbol{\Omega})^{-1} \boldsymbol{e}_t. \tag{28}$$

It is to be noted that our objective function here is different from that in Gallant and Golub (1984), due to different assumptions concerning the error vector. The specification of the objective function as a log likelihood function enables a comparison of the constrained maximum likelihood approach (that is, the constrained optimization) with the Bayesian approach where we will use non-informative priors for the parameters, thus rendering parameter estimates to be mainly determined by the likelihood function.

We first estimate an unconstrained version of (28) and check the theoretical regularity conditions of positivity, concavity, and monotonicity. In cases where positivity, concavity, and monotonicity are not satisfied, we then use the NPSOL nonlinear programming program to maximize $\ln L\left(\boldsymbol{w} \middle| \widetilde{\boldsymbol{\beta}}, \boldsymbol{\Omega}\right)$ with the positivity, concavity and monotonicity conditions imposed. While we follow Gallant and Golub (1984) and use nonlinear constrained optimization to impose concavity, we do not do it by constructing their submatrix $\boldsymbol{K}_{22}$ using a Householder transformation and then deriving an indicator function for the smallest eigenvalue of $\boldsymbol{K}_{22}$ and its derivative. Instead, we work directly with matrix $\boldsymbol{G}$ by restricting its eigenvalues to be nonpositive. This is because a necessary and sufficient condition for negative semidefiniteness of $\boldsymbol{G}$ is that all its eigenvalues are nonpositive — see, for example, Morey

(1986). Compared with the Gallant and Golub (1984) approach where a reduced matrix $\boldsymbol{K}_{22}$ is sought, the direct restriction of the eigenvalues of $\boldsymbol{G}$ to be nonpositive seems to be more appealing.

Thus, our constrained optimization problem can be written as follows

$$\max_{\boldsymbol{\theta}} \ln L\left(\boldsymbol{w} \left|\boldsymbol{\theta}\right.\right),$$

subject to

$$C_t\left(\boldsymbol{p}, y, t\right) > 0,\ \ t = 1, ..., T \tag{29}$$

$$\varphi_{it}\left(\boldsymbol{p}, y, t, \boldsymbol{\theta}\right) < 0, \quad i = 1, \cdots, M,\ \ t = 1, ..., T \tag{30}$$

$$s_{it}\left(\boldsymbol{p}, y, t, \widetilde{\boldsymbol{\beta}}\right) > 0, \quad i = 1, \cdots, M,\ \ t = 1, ..., T \tag{31}$$

where $\boldsymbol{\theta}$ is the parameter vector to be estimated, $\boldsymbol{\theta} = (\widetilde{\boldsymbol{\beta}}, \boldsymbol{\Omega})$, and $\varphi_i\left(\boldsymbol{p}, y, t, \boldsymbol{\theta}\right)$, $i = 1, \cdots, M$, are the M estimated eigenvalues of the $\boldsymbol{G}$ matrix at time t. (30) can be further simplified by requiring the largest eigenvalue to be nonpositive.

The constrained optimization approach is capable of imposing curvature locally, regionally, and fully, in that we can evaluate $\boldsymbol{G}$ at a single data point, over a neighborhood of interest, or at every sample observation. In addition, it is also capable of imposing concavity globally. To do this, we just need to maximize the log likelihood function (28) subject to the eigenvalues of the matrix $\boldsymbol{B}$ being nonpositive. Alternatively, this can also be done by maximizing the log likelihood function (28) subject to the constraint indicator of $-\boldsymbol{B}$ being zero, or subject to all determinants of the leading principal minors of $\boldsymbol{B}$ being nonpositive. In this paper, we use the former method.

In contrast to the Cholesky factorization approach which is only capable of imposing local monotonicity and local positivity only, the constrained optimization approach is capable of imposing monotonicity and positivity locally, regionally, and fully. For the case of monotonicity, this can be done by adding the constraints, $s_{it} > 0$, $(i = 1, \cdots, M,\ t = 1, ..., T)$, at the single reference point, over the neighborhood of interest, or at every data point. For the case of the positivity, this can be done by adding the constraints, $C_t > 0$ $(t = 1, ..., T)$, at the single reference point, over the neighborhood of interest, or at every data point. Compared with the Cholesky factorization approach, the superiority of the constrained optimization approach in imposing regional and pointwise monotonicity and positivity is due to its ability in manipulating both parameter and data spaces.

The constrained optimization approach is also incapable of imposing global monotonicity. As can be seen from (26), global monotonicity can be imposed by setting $\beta_i > 0$, $\beta_{ij} > 0$, and $\beta_{it} > 0$ if $\ln\left(p_j/p_M\right) > 0$ and $t > 0$. However, there are two reasons for which the imposition of global monotonicity may not work. First, $\ln\left(p_j/p_M\right)$ can be negative if $p_j < p_M$. Second, $\beta_{ij} > 0$ contradicts the concavity condition, which requires the principal minors that are of an odd-numbered order be non-positive and the principal minors that are of an even-numbered order be non-negative — for the sufficient and necessary conditions for concavity, see Morey (1986).

Similarly, the constrained optimization approach is also incapable of imposing global positivity. As can be seen from (25), global positivity can be imposed by setting $\beta_0 > 0$, $\beta_t > 0$, $\beta_i > 0$, $\beta_{ij} > 0$, $\beta_{it} > 0$, and $\beta_{tt} > 0$, if $\ln y > 0$, $\ln(p_j/p_M) > 0$ and $t > 0$. However, for the same reasons as discussed in the previous paragraph, the imposition of global positivity may not work.

4.3 The Bayesian Approach

Recently, numerical integration simulation techniques, such as Gibbs sampling, introduced by Geman and Geman (1984), and the Metropolis-Hastings algorithm, due to early work by Metropolis *et al.* (1953) and Hastings (1970), have made Bayesian estimation of the seemingly unrelated regression (SUR) model more convenient and accessible. Terrell (1996), Koop *et al.*, (1997), and Griffiths *et al.* (2000) extend this literature by incorporating positivity, monotonicity and concavity conditions into the estimation of flexible functional forms. This subsection draws mainly on the discussion of Bayesian estimation of cost and share equation system by Griffiths *et al.* (2000).

As in the constrained optimization approach, we impose linear homogeneity by normalizing cost and the input prices by the last input price, p_M, to obtain (25) and (26). To impose the cross-equation equality restrictions on the parameters of equations (25)-(26), we first treat all the parameters as if they were 'unrestricted.' In particular, we rewrite the $M-1$ share equations at time t $(t = 1, ..., T)$ in (26), after appending an error component, as

$$w_i = \boldsymbol{X}_i \boldsymbol{\alpha}_i + e_i, \quad i = 1, \cdots, M-1, \tag{32}$$

where w_i, $\boldsymbol{X}_i$, and e_i $(i = 1, \cdots, M-1)$ are defined as in equation (27). The vectors $\boldsymbol{\alpha}_i = (\alpha_{i1}, \cdots, \alpha_{i,M+1})'$ for $i = 1, \cdots, M-1$ are the corresponding vectors of unrestricted parameters to be estimated. Similarly, we rewrite equation (25) at time t, after appending an error component, as

$$w_M = \boldsymbol{X}_M \boldsymbol{\alpha}_M + e_M, \tag{33}$$

where w_M, $\boldsymbol{X}_M$, and e_M are defined as in (27) and $\boldsymbol{\alpha}_M = \left(\alpha_{M1}, \cdots, \alpha_{M,(M^2+3M+2)/2}\right)'$ is the corresponding vector of unrestricted parameters to be estimated. Thus, instead of having $(M^2 + 3M + 2)/2$ parameters in (25) and (26), we now have $3M(M + 1)/2$ parameters in (32) and (33). The empirical model, as defined by (32) and (33), can be written compactly as

$$\boldsymbol{w}_t = \boldsymbol{X}_t \boldsymbol{\alpha} + \boldsymbol{e}_t, \tag{34}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_M)'$.

It should be noted that the $\boldsymbol{\alpha}_j$ $(j = 1, \cdots, M)$ vectors have many elements in common. In the case with $M = 4$, for example, there are 15 cross-equation equality restrictions, given as follows: $\alpha_{11} = \alpha_{42}$, $\alpha_{12} = \alpha_{46}$, $\alpha_{13} = \alpha_{47}$, $\alpha_{14} = \alpha_{48}$, $\alpha_{15} = \alpha_{4,12}$, $\alpha_{21} = \alpha_{43}$, $\alpha_{22} = \alpha_{47}$, $\alpha_{23} = \alpha_{49}$, $\alpha_{24} = \alpha_{4,10}$, $\alpha_{25} = \alpha_{4,13}$, $\alpha_{31} = \alpha_{44}$, $\alpha_{32} = \alpha_{48}$, $\alpha_{33} = \alpha_{4,10}$, $\alpha_{34} = \alpha_{4,11}$, and $\alpha_{35} = \alpha_{4,14}$. Therefore, among the 30 parameters, 15 [those in the cost equation (33)] are free and the rest [those in the share equations (32)] have to be obtained from these free parameters. The $(M^2 - 1)$ restrictions implied by (32) and (33) can be written more compactly as

$$\boldsymbol{R}\boldsymbol{\alpha} = \boldsymbol{r}, \tag{35}$$

where $\boldsymbol{R}$ and $\boldsymbol{r}$ are known matrices of order $(M^2 - 1) \times 3M(M + 1)/2$ and $(M^2 - 1) \times 1$, respectively.

To impose the $(M^2 - 1)$ cross-equation restrictions, we re-order the elements in $\boldsymbol{\alpha}$ and $\boldsymbol{R}$ such that (35) can be rewritten as

$$\boldsymbol{R}\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_2 \end{bmatrix} \begin{bmatrix} \widetilde{\gamma} \\ \gamma \end{bmatrix} = \boldsymbol{r}, \tag{36}$$

where $\boldsymbol{R}_1$ is a $(M^2-1)\times(M^2-1)$ matrix, and $\boldsymbol{R}_2$ a $(M^2-1)\times(M^2+3M+2)/2$ matrix. The vector $\boldsymbol{\gamma}$ contains the subset of $(M^2 + 3M + 2)/2$ free parameters, and the vector $\widetilde{\boldsymbol{\gamma}}$ contains the remaining parameters in $\boldsymbol{\alpha}$ which will be recovered using $\boldsymbol{\gamma}$. Correspondingly, we can reorder the columns of $\boldsymbol{X}$ and partition it so that the linear SUR model can be written as

$$\begin{aligned} \boldsymbol{w}_t &= \boldsymbol{X}_t\boldsymbol{\alpha} + \boldsymbol{e}_t \\ &= \begin{bmatrix} \boldsymbol{X}_{1t} & \boldsymbol{X}_{2t} \end{bmatrix} \begin{bmatrix} \widetilde{\gamma} \\ \gamma \end{bmatrix} + \boldsymbol{e}_t. \end{aligned} \tag{37}$$

From (36), we can solve for $\widetilde{\boldsymbol{\gamma}}$ as

$$\widetilde{\gamma} = \boldsymbol{R}_1^{-1} \left[\boldsymbol{r} - \boldsymbol{R}_2\gamma\right]. \tag{38}$$

Substituting (38) into (37) and rearranging yields

$$\boldsymbol{w}_t - \boldsymbol{X}_{1t}\boldsymbol{R}_1^{-1}\boldsymbol{r} = \left(\boldsymbol{X}_{2t} - \boldsymbol{X}_{1t}\boldsymbol{R}_1^{-1}\boldsymbol{R}_2\right)\gamma + \boldsymbol{e}_t$$

or

$$\boldsymbol{z}_t = \boldsymbol{Z}_t\gamma + \boldsymbol{e}_t, \tag{39}$$

where $\boldsymbol{z}_t = \boldsymbol{w}_t - \boldsymbol{X}_{1t}\boldsymbol{R}_1^{-1}\boldsymbol{r}$ and $\boldsymbol{Z}_t = \boldsymbol{X}_{2t} - \boldsymbol{X}_{1t}\boldsymbol{R}_1^{-1}\boldsymbol{R}_2$ represent new sets of observations, and $\boldsymbol{e}_t$ is defined as above.

The formulation of our empirical model as an unrestricted SUR model (39) is convenient for Bayesian analysis. The Bayes Theorem in our particular case can be restated as

$$f\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\,|\boldsymbol{z}\right)\propto L\left(\boldsymbol{z}\,|\boldsymbol{\gamma},\boldsymbol{\Omega}\right)p\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\right), \tag{40}$$

where $f\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\,|\boldsymbol{z}\right)$ is the posterior joint density function for $(\boldsymbol{\gamma},\boldsymbol{\Omega})$ given $\boldsymbol{z}$, $L\left(\boldsymbol{z}\,|\boldsymbol{\gamma},\boldsymbol{\Omega}\right)$ is the likelihood function, which summarizes all the sample information, and $p\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\right)$ is the joint prior density function for $(\boldsymbol{\gamma},\boldsymbol{\Omega})$.

Under the assumption that $\boldsymbol{e}_t$ is iid normal, the likelihood function in (40) can be shown to be

$$L\left(\boldsymbol{z}\,|\boldsymbol{\gamma},\boldsymbol{\Omega}\right)\propto|\boldsymbol{\Omega}|^{-\frac{T}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Omega}^{-1}\right)\right], \tag{41}$$

where $\boldsymbol{A}$ is a $M\times M$ symmetric matrix with $a_{ij}=(\boldsymbol{z}_i-\boldsymbol{Z}_i\boldsymbol{\gamma})'(\boldsymbol{z}_j-\boldsymbol{Z}_j\boldsymbol{\gamma})$ $i=1,2,...,M$.

The Bayesian model in (40) requires that we choose priors for the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Omega}$. We choose a flat (constant) prior for $\boldsymbol{\gamma}$,

$$p\left(\boldsymbol{\gamma}\right)\propto I\left(\boldsymbol{\gamma}\in R_j\right), \tag{42}$$

where $I\left(\cdot\right)$ is an indicator function which takes the value one if the argument is true and zero otherwise, and R_j is the set of permissible parameter values when no constraints ($j=0$) and the positivity, monotonicity and curvature constraints ($j=1$) must be satisfied. With the constant equal to an indicator function, our particular flat prior for $\boldsymbol{\gamma}$ allows us to slice away the portion of the posterior density that violates positivity, monotonicity and curvature of the translog cost function.

We adopt the following prior for $\boldsymbol{\Omega}$

$$p(\boldsymbol{\Omega})\propto|\boldsymbol{\Omega}|^{-(M+1)/2}, \tag{43}$$

which is the limiting form of the inverted Wishart density — see Griffiths *et al.* (2000).

Combining the priors (42) and (43) gives the joint prior probability density function

$$p\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\right)\propto|\boldsymbol{\Omega}|^{-(T+1)/2}I\left(\boldsymbol{\gamma}\in R_j\right). \tag{44}$$

Combining the likelihood function in (41) and the joint prior distribution (44), yields the posterior joint density function

$$f\left(\boldsymbol{\gamma},\boldsymbol{\Omega}\,|\boldsymbol{z}\right)\propto|\boldsymbol{\Omega}|^{-(T+M+1)/2}I\left(\boldsymbol{\gamma}\in R_j\right)\exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Omega}^{-1}\right)\right], \tag{45}$$

from which we can derive full conditional posterior pdfs for $\boldsymbol{\gamma}$ and $\boldsymbol{\Omega}$. We then use the Gibbs sampling algorithm which draws from the joint posterior density by sampling from a series of conditional posteriors. Essentially, Gibbs sampling involves taking sequential random draws from full conditional posterior distributions. Under very mild assumptions [see, for example, Tierney (1994)], these draws converge to draws from the joint posterior. Once draws from the joint distribution have been obtained, any posterior feature of interest can be calculated.

The full conditional posterior distributions for $\boldsymbol{\gamma}$ can be obtained from (45) by treating $\boldsymbol{\Omega}$ as a constant

$$p\left(\boldsymbol{\gamma}\,|\boldsymbol{z},\boldsymbol{\Omega}\right) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}\right)'\boldsymbol{V}\left(\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}\right)\right] I\left(\boldsymbol{\gamma}\in R_j\right), \tag{46}$$

where $\widehat{\boldsymbol{\gamma}}$ is the GLS estimator for $\boldsymbol{\gamma}$ in (39)

$$\widehat{\boldsymbol{\gamma}} = \left[\boldsymbol{Z}'\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\iota}_T\right)\boldsymbol{Z}\right]^{-1}\boldsymbol{Z}'\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\iota}_T\right)\boldsymbol{z},$$

$\boldsymbol{V} = \boldsymbol{Z}'\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\iota}_T\right)\boldsymbol{Z}$, and $\boldsymbol{\iota}_T$ is a $T\times T$ identity matrix. The full conditional posterior distributions for $\boldsymbol{\Omega}$ can be obtained from (45) by treating $\boldsymbol{\gamma}$ as a constant

$$p\left(\boldsymbol{\Omega}\,|\boldsymbol{z},\boldsymbol{\gamma}\right) \propto |\boldsymbol{\Omega}|^{-(T+M+2)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Omega}^{-1}\right)\right]. \tag{47}$$

Thus $p\left(\boldsymbol{\Omega}\,|\boldsymbol{z},\boldsymbol{\gamma}\right)$ is an inverted Wishart density function with parameters $\boldsymbol{A}$, T, and M.

The Gibbs sampler for Bayesian estimation without positivity, monotonicity, and curvature constraints can be implemented by setting $I\left(\boldsymbol{\gamma}\in R_0\right)$ in (46) equal to one and then drawing sequentially from the above conditional posteriors in (46) and (47). Sampling from these two conditional posteriors is straightforward.

There are two approaches in this literature which can be used to impose positivity, monotonicity, and curvature conditions: the accept-reject algorithm [see Terrell (1996)] and the random-walk Metropolis Hastings (M-H) algorithm, proposed by Griffiths *et al.* (2000). The accept-reject algorithm, however, has been criticized for its inefficiency in that it needs to generate an extremely large number of candidate draws before finding one that is acceptable — see Griffiths *et al.* (2000) for more details. In this paper, we impose positivity, monotonicity, and curvature condition through the Metropolis Hastings algorithm, which in our particular case proceeds iteratively as follows – see Griffiths *et al.* (2000) for more details:

- *Step* 1: Start with an initial value $\boldsymbol{\gamma}^j$ satisfying both the positivity, monotonicity and curvature constraints. Let j denote the state of $\boldsymbol{\gamma}$ and set $j=1$. Given the complexity of the cost and share equation system, the search for the initial value which satisfies the positivity, monotonicity and curvature constraints is not a trivial issue in practice. A convenient way, as done in this paper, is to use the estimated $\widetilde{\boldsymbol{\beta}}$ from the constrained optimization model outlined above.

- *Step* 2: Using the current value $\boldsymbol{\gamma}^j$, sample a candidate point $\boldsymbol{\gamma}^c$ from a symmetric proposal density $q(\boldsymbol{\gamma}^c,\boldsymbol{\gamma}^j)$, which is the probability of returning a value of $\boldsymbol{\gamma}^c$ given a previous value of $\boldsymbol{\gamma}^j$.

- *Step* 3: Evaluate the positivity, monotonicity and curvature constraints at the specified data points using the candidate value, $\boldsymbol{\gamma}^c$. If any constraints are violated, set $\alpha(\boldsymbol{\gamma}^j,\boldsymbol{\gamma}^c) = 0$ and go to Step 5.

- *Step* 4: Calculate $\alpha(\boldsymbol{\gamma}^j,\boldsymbol{\gamma}^c) = \min\left[h(\boldsymbol{\gamma}^c)/h(\boldsymbol{\gamma}^j), 1\right]$ where $h(\boldsymbol{\gamma}) = |\boldsymbol{A}|^{-T/2} I(\boldsymbol{\gamma} \in R_j)$ is the kernel of $p(\boldsymbol{\gamma}\,|\boldsymbol{z})$, which is obtained by integrating $\boldsymbol{\Omega}$ out of the joint posterior (45).

- *Step* 5: Generate an independent uniform random variable u from the interval $[0, 1]$.

- *Step* 6: Set $\boldsymbol{\gamma}^{j+1} = \boldsymbol{\gamma}^c$ if $\boldsymbol{u} < \alpha(\boldsymbol{\gamma}^j,\boldsymbol{\gamma}^c)$ and $\boldsymbol{\gamma}^j$ otherwise.

- *Step* 7: Set $j = j + 1$ and return to Step 2.

The algorithm works best if the proposal density matches the shape of the target distribution. Therefore, the proposal density is chosen to be a multivariate normal with mean equal to the current state $\boldsymbol{\gamma}^j$ and covariance matrix equal to $\left(\boldsymbol{Z}'\left(\widehat{\boldsymbol{\Omega}}^{-1} \otimes \boldsymbol{\iota}_T\right)\boldsymbol{Z}\right)^{-1}$ [where $\widehat{\boldsymbol{\Omega}}$ is an estimator of $\boldsymbol{\Omega}$ constructed from using OLS residuals] multiplied by a tuning parameter. The tuning parameter can be used to adjust the acceptance rate. The optimal acceptance rate (i.e., the one which minimizes the autocorrelations across the sample values) has been shown to lie within the range between 0.45 (in one-dimensional problems) and approximately 0.23 (as the number of dimensions becomes infinitely large). In our empirical work, we choose the tuning parameter so that the acceptance rate lies within this range.

The Bayesian approach is capable of imposing curvature locally, regionally, and fully. In particular, the curvature constraint mentioned in step 3 can be evaluated by using eigenvalues, a constraint indicator, or the determinants of the leading principal minors of the matrix $\boldsymbol{G}$. The Bayesian approach is also capable of imposing concavity globally. To do this, we just need to evaluate the curvature constraint mentioned in step 3 by using eigenvalues or the determinants of the leading principal minors of $\boldsymbol{B}$, instead of $\boldsymbol{G}$.

Like with the optimization approach, with the Bayesian approach we can impose the monotonicity (positivity) condition locally, regional, and fully. In particular, we can evaluate s_i (C, i.e., total cost) in (6) at a single data point, over a neighborhood of interest, or at every sample observation, every time the candidate value $\boldsymbol{\beta}^c$ is drawn. If any s_i $(i = 1, \cdots, M)$ is negative at data point(s) of interest, we set $\alpha(\boldsymbol{\gamma}^j,\boldsymbol{\gamma}^c) = 0$ and go to Step 5. However, for the same reason as in the case of the constrained optimization approach, we cannot impose global monotonicity (or global positivity) with the Bayesian approach.

An important difference between the Bayesian and optimization approaches is that, in the Bayesian approach, the constraints will be satisfied for all parameters values where the posterior density is nonzero. That means that all values in probability intervals for parameters or elasticities will be consistent with the constraints. In the sampling theory approach only the estimates, not necessarily all values within confidence intervals, will satisfy the constraints.

5 Application to KLEM Data

5.1 Data

We use the Feng and Serletis (2008) annual KLEM (capital, labor, energy, and intermediate materials) data for total manufacturing in the United States over the period from 1953 to 2001, obtained from the website of the U.S. Bureau of Labor Statistics (BLS), at www.bls.gov/data/home.htm. This data set is similar to the Berndt and Wood (1975) classical annual data set for the U.S. manufacturing sector over the period 1947-1971, which has been used in many influential papers in this literature — see, for example, Gallant and Golub (1984), Diewert and Wales (1987), and Terrell (1996). The data consists of price and quantity indices for one output and four inputs (capital, labor, energy, and materials). All the price series have been normalized to one in 1953 and the quantity indices for output, capital, labor, energy, materials, and purchased business services have been obtained by dividing value of production or factor costs by the corresponding normalized price index. It should be noted that we constructed the price and quantity indices for intermediate materials as subaggregates over the two components, materials and purchased business services, using the Fisher ideal index.

5.2 Empirical Results

We first use the LSQ command in TSP, which is actually based on the maximum likelihood method, to estimate both the unconstrained translog model and the constrained one with curvature imposed using the Cholesky factorization approach. We then use the TOMLAB/NPSOL tool box with MATLAB for both the unconstrained and the constrained optimization. Since the objective of the optimization method is the log likelihood function, the unconstrained and constrained optimization methods can be also regarded as unconstrained and constrained maximum likelihood (ML) methods. Finally, we use GAUSS for both unconstrained and constrained Bayesian estimation.

5.2.1 Theoretical Regularity Tests

Table 1 summarizes the results from the three different estimation methods (i.e., maximum likelihood estimation using LSQ, optimization, and the Bayesian approach) in terms of parameter estimates, log likelihood values, and theoretical regularity (positivity, monotonicity, and curvature) violations when regularity conditions are not imposed. As can seen from the bottom of Table 1, positivity and monotonicity are satisfied at all data points, whereas concavity is violated at all data points.

With regards to the statistical inferences, standard errors are provided for the unconstrained maximum likelihood estimation (using the LSQ command in TSP) and standard deviations of the MCMC samples are provided for the Bayesian method of estimation. As in Gallant and Golub (1984), standard errors are not reported for the unconstrained optimization. Of course, a parametric bootstrap method could be used, as in Gallant and Golub (1984), to compute standard errors or confidence intervals for the parameter estimates. This involves the use of Monte Carlo methods to obtain a reliable estimate of the sampling distribution, by generating a large enough sample from the distribution of the constrained estimator.

Clearly, the results in Table 1 indicate that the parameter estimates, log likelihood value, and theoretical regularity violations are identical under the unconstrained maximum likelihood method of estimation (by using the LSQ command in TSP) and the unconstrained optimization. This is not surprising in that the unconstrained optimization is essentially unconstrained maximum likelihood estimation in our particular case, as noted above. Moreover, the parameter estimates from the unconstrained Bayesian estimation are very close to those from maximum likelihood estimation (columns 1 and 2), implying that the results from the Bayesian method are mainly determined by the likelihood function, due to the noninformative prior for $\boldsymbol{\gamma}$. It is to be noted that with the maximum likelihood method of estimation (by using the LSQ command in TSP) and unconstrained optimization we evaluate the theoretical regularity conditions using the point estimates of the parameters. With the Bayesian method of estimation, regularity tests are performed by evaluating the posterior means of s_i and the eigenvalues of $\boldsymbol{G}$.

5.2.2 Effectiveness in Imposing Curvature

Because curvature is not satisfied under any of the three methods of estimation, we follow the procedures discussed in Section 4 and impose curvature. In doing so, we make a distinction between 'local concavity imposition' (at a single point in the data), 'regional concavity imposition' (over a neighborhood of data points in the sample), and 'pointwise concavity imposition' (at every data point in the sample). Moreover, given that concavity is a property of conditional inputs satisfying the law of demand, we also impose 'global concavity,' occurring at all possible, positive input prices.

Table 2 summarizes the results from the Cholesky factorization approach using maximum likelihood estimation. With this approach we cannot impose regional or pointwise concavity; we can only impose local concavity (at a single point) and global concavity. Regarding local concavity, as noted by Ryan and Wales (2000), the ability of locally flexible models to satisfy curvature at sample observations other than the point of reference, depends on the choice of the reference point. Thus, we estimate the model 49 times (a number of times equal to the number of observations) and report results for the best reference point (best in the sense of satisfying the curvature conditions at the largest number of observations). The best reference point is 1970 ($t = 28$). While Ryan and Wales (2000) found that imposing concavity at the mean yields satisfaction of concavity at all points in their data set, our finding in terms of curvature violations when the curvature conditions are imposed locally is disappointing. In particular, imposition of curvature locally (at $t = 28$) simply reduces the number of curvature violations from 49 to 6 (see column 1 of Table 2).

Regarding global concavity, as mentioned earlier the translog will satisfy concavity at all possible, positive input prices, assuming positive shares, if the $\boldsymbol{B}$ matrix is negative semidefinite. To impose this condition, we exploit the key idea behind the Cholesky factorization approach and set $\boldsymbol{B} = -\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{K}$ is a lower triangular matrix. After replacing the elements of $\boldsymbol{B}$ with those of $\boldsymbol{K}$, (8) is estimated and the elements of $\boldsymbol{B}$ are recovered from the restriction $\boldsymbol{B} = -\boldsymbol{K}\boldsymbol{K}'$. As can be seen in column 2 of Table 2, imposition of global concavity eliminates all the curvature violations, without inducing positivity or monotonicity violations. We also note that the predominance of zeros for the β_{ij} suggests the translog reduces to a Cobb-Douglas when global concavity is imposed.

In Table 3 we present the results from the constrained optimization approach. With this approach we impose concavity regionally, fully, and globally; we do not impose concavity locally, because we cannot find a point in the data that leads to satisfaction of theoretical regularity at all points in the data set. In imposing concavity regionally, we find that there are more than one regions within the data set that can yield to satisfaction of theoretical regularity at all data points in the data set. However, the log likelihood values are different for these different regions and in column 1 of Table 3 we report results only for the best neighborhood (best in the sense of satisfying the theoretical regularity conditions with the largest log likelihood value), which is data points 16-20. Column 2 reports the results when concavity is imposed fully (that is, at every data point). It is interesting to note here that the imposition of concavity regionally and the imposition of concavity fully result in the same parameter estimates and log likelihood values. This is so, because only the constraints on data points 16-20 are binding, and the constraints on the rest of the data points are slack. Finally, as noted earlier, in imposing concavity globally (in column 3 of Table 3), we maximize the log likelihood function (28) subject to the eigenvalues of $\boldsymbol{B}$ being non-positive (that is, subject to $\boldsymbol{B}$ being negative semidefinite). Clearly, positivity, monotonicity, and curvature are all satisfied at all data points when curvature is imposed regionally, fully, and globally using constrained optimization.

In Table 4 we present the results (together with 90% posterior credible intervals) from constrained Bayesian estimation. We impose concavity at three different levels — locally, fully, and globally. Columns 1 and 2 report the results when concavity is imposed locally. Again, we report the results for the best point in the data set (best in the sense of satisfying all three theoretical regularity conditions at every data point with the largest log likelihood value). The best data point is $t = 17$; in this case there is no need to impose curvature regionally. Columns 3 and 4 report the results when concavity is imposed fully (at every data point in the sample). In imposing global concavity with the Bayesian approach (see columns 5 and 6), we still need $\boldsymbol{B}$ to be negative semidefinite. In particular, we evaluate the eigenvalues of $\boldsymbol{B}$ every time the candidate value $\boldsymbol{\gamma}^c$ is drawn using the Metropolis Hastings algorithm, discussed earlier in Section 4. If any of the eigenvalues of $\boldsymbol{B}$ obtained from $\boldsymbol{\gamma}^c$ are positive, we set $\alpha(\boldsymbol{\gamma}^j,\boldsymbol{\gamma}^c) = 0$ and go to Step 5. It is clear that theoretical regularity is satisfied at all data points when concavity is imposed locally, fully, and globally, using Bayesian methods.

5.2.3 The Effect on Flexibility

The imposition of the theoretical regularity restrictions almost always compromises the flexibility of functional forms. Thus, an ideal method will be one that can guarantee inference consistent with theory, but without compromising much of the flexibility of the functional form. In this section, we evaluate the effectiveness in maintaining flexibility by examining the effect of the imposition of concavity on the estimated parameters and price elasticities.

A preliminary examination of the effects of the imposition of concavity on the flexibility of the functional form can be achieved by comparing the log likelihood values in Tables 1-4. In the case, for example, of LSQ estimation and imposition of the curvature restrictions using reparameterization based on the Cholesky factorization approach, the log likelihood value decreases from 713.484 in the unrestricted case to 561.627 when curvature is imposed globally; the large decline in the log likelihood values when curvature is imposed locally is due to the normalization of the input prices and output at a different data point rather than the imposition of local curvature. The large effect of imposition of global curvature can also be seen from the large changes in the estimated parameters. In particular, the elements of the matrix $\boldsymbol{B}$ are forced to be close to zero by the imposition of global curvature. As with the LSQ method of estimation, the imposition of global concavity using constrained optimization and the Bayesian approach also has large effects on the value of the log likelihood function and the estimated parameters, forcing the elements of $\boldsymbol{B}$ to be close to zero. Thus, as far as log likelihood values and changes in estimated parameters are concerned, the imposition of local, regional, or pointwise concavity using constrained optimization or Bayesian methods is the preferred approach, due to the small loss of flexibility of the functional form.

Further insight regarding the effects of imposing theoretical regularity on the flexibility of the functional from can be gained by examining the effect on the estimates of the price elasticities. These elasticities, calculated as

$$\eta_{ij} = \frac{\beta_{ij}}{s_i} + s_j - \delta_{ij}, \tag{48}$$

where δ_{ij} is defined as in (9), are reported in Table 5 for 1953 (the beginning of the sample period) and in Table 6 for 2001 (the last year of the sample period).

When curvature is imposed locally, regionally, or fully, using optimization and the Bayesian approach, all input price elasticities are reasonably close to those under the unrestricted case. For example, the four own price elasticities, η_{11}, η_{22}, η_{33}, and η_{44}, obtained from the optimization approach at the last year of the sample period (see Table 6) have changed from -0.214, 0.022, -0.481, and -0.090 in the unrestricted case to -0.215, -0.068, -0.591, and -0.154, respectively, when concavity is imposed regionally at data points 16-20 or fully (at all data points). This is consistent with the small reduction in the log likelihood value from 713.484 in the unconstrained optimization (see Table 1) to 706.769 when curvature is imposed regionally or fully (see Table 3). When, however, curvature is imposed globally, the own-price elasticities become much larger in absolute value; this is true for all methods of estimation — LSQ, optimization, and the Bayesian method. For example, under the optimization approach, η_{11}, η_{22}, η_{33}, and η_{44} change to -0.821, -0.618, -0.965, and -0.596, respectively, when global concavity is imposed. In other words, $\boldsymbol{G}$ (or $\widetilde{\boldsymbol{G}}$) is made too negative semidefinite by the imposition of global concavity.

The effect on the price elasticity estimates of the imposition of concavity can also be seen from the marginal posterior pdfs for the price elasticities under the Bayesian method of estimation. In Figure 1, for example, we show the marginal posterior pdfs for η_{22} under unrestricted Bayesian estimation as well as for the cases when curvature is imposed locally (at $t = 17$), fully (at every point), and globally (at every possible, positive input price). Clearly, the imposition of local concavity shifts the posterior mean of η_{22} to the left from a positive value (0.014) to a small negative value of -0.076. The imposition of pointwise concavity further shifts the posterior mean of η_{22} to the left to a negative value with a slightly larger magnitude of -0.099. However, the posterior mean of η_{22} is shifted too much to the left to -0.661, when global concavity is imposed. In other words, the sensitivity of the demand for capital to its own price is exaggerated when global concavity is imposed.

Moreover, for all three methods — Cholesky factorization, optimization, and the Bayesian methodology — the imposition of global concavity rules out most possibilities of a complementarity relationship among the inputs. The imposition, however, of local, regional, or pointwise concavity preserves most of the substitutability/complementarity relationships among the inputs relative to the unconstrained model. In particular, all cross-price elasticities (except η_{34}) are positive under both optimization and Bayesian estimation when global concavity is imposed, meaning that no pairs of inputs (except energy and materials) are complements. This can be seen by examining equation (48). As already noted, when global concavity is imposed, the elements of $\boldsymbol{B}$ are forced to be close to zero, which in turn reduces equation (48) to

$$\sigma_{ij} \approx s_j,$$

since δ_{ij} is also zero when $i \neq j$. Since s_j (the cost share for input j) is always positive (when monotonicity holds), σ_{ij} will also be positive in most cases, meaning most pairs of inputs are forced to be substitutes to each other when global concavity is imposed. In contrast, out of the twelve cross price elasticities, only four change signs by the imposition of local, regional, or pointwise concavity, under either optimization or Bayesian estimation — those being η_{23}, η_{24}, η_{32}, and η_{42}. Hence, the imposition of local, regional, or pointwise concavity in addition to assuring theoretical regularity at every data point, it also preserves the substitutability/complementarity relationships among the inputs relative to the unrestricted model.

6 Conclusion

Most published studies that use flexible functional forms have ignored the theoretical regularity conditions implied by microeconomic theories. Moreover, even a few studies have checked and/or imposed regularity conditions, most of them equate curvature alone with regularity, thus ignoring or minimizing the importance of monotonicity. We argue that without satisfaction of all three theoretical regularity conditions (positivity, monotonicity, and curvature), the resulting inferences are questionable, since violations of regularity conditions invalidate the duality theory on which economic functions are based. We believe that unless regularity is attained by luck, flexible functional forms should always be estimated subject to regularity.

In this paper we provide an empirical comparison of three different methods of imposing theoretical regularity on flexible cost functions — Cholesky factorization, constrained optimization, and Bayesian methodology. In doing so, we chose the translog cost function, due to its widespread applications in both the traditional factor demand system literature and the more recent firm efficiency literature.

With the Cholesky factorization approach, regularity conditions are imposed through the decomposition of a symmetric, positive (or negative) definite matrix, expressed in terms of parameters only, into the product of a lower triangular matrix and its conjugate transpose. If a regularity condition can not reduce to a symmetric, positive (or negative) definite matrix, expressed in terms of parameters only, this approach will be useless. In the case of translog cost function, this approach is capable of imposing local (at a single point) and global (at all possible, positive input prices) curvature, as well as local positivity and monotonicity. However, it is incapable of imposing regional (at a set of data points) or pointwise (at every sample observation) curvature. In addition, it is also incapable of imposing regional, pointwise, or global positivity and monotonicity. This is because the imposition of these regularity conditions not only involves the manipulations of parameter space, but also involves the manipulations of data space, which is impossible with this approach.

With the constrained optimization approach, the imposition of theoretical regularity conditions (positivity, monotonicity and curvature) involves specifying an objective function and some inequality constraints corresponding to the monotonicity and curvature restrictions. More specifically, with this approach, regularity conditions are imposed by maximizing (or minimizing) the value of the objective function within the parameter space defined by the constraints. This approach is workable as long as the inequality constraints can be explicitly specified. In the case of translog cost function, this approach is capable of imposing local, regional, pointwise, and global curvature, as well as local, regional, and pointwise positivity and monotonicity. However, with this approach, we can not impose global positivity and monotonicity, since these regularity conditions can not be explicitly specified. It should be noted that a limitation associated with this approach is that statistical inferences usually can not be obtained directly, and one thus has to resort to resampling methods (e.g., bootrap methods) to obtain them.

As for the Bayesian approach, a Metropolis-Hastings algorithm is used within the Gibbs sampler to impose theoretical regularity conditions and simulate observations from truncated pdfs. Like the constrained optimization approach, this approach is workable as long as the inequality constraints can be explicitly specified. In the case of translog cost function, this approach is capable of imposing local, regional, pointwise, and global curvature, as well as local, regional, and pointwise positivity and monotonicity. However, for the same reason as in the case of the constrained optimization approach, we can not impose global positivity and monotonicity with this approach. Compared with the constrained optimization approach, the Bayesian approach has one major advantage, that is, it is easy to obtain statistical inferences for the parameters and any measures (e.g., elasticities and productivity measures) which can be expressed as a function of the parameters.

We applied the above three methods to the annual KLEM (capital, labor, energy, and intermediate materials) data for total manufacturing in the United States over the period from 1953 to 2001. We find that positivity and monotonicity are satisfied automatically. As for the imposition of curvature conditions, we find that while the Cholesky factorization approach can successfully impose curvature at the reference point, it does not guarantee curvature satisfaction at every sample observation. For the case of the constrained optimization approach, the imposition of local curvature does not assure theoretical regularity at every sample observation too; however, the imposition of regional or pointwise curvature leads to the satisfaction of curvature conditions at every sample observation without significant loss of flexibility. As for the Bayesian approach, the imposition of local, regional, and pointwise curvature leads to the satisfaction of curvature conditions at every sample observation without significant loss of flexibility. Finally, we find that the imposition of global curvature, irrespective of the method of imposition, forces the elements of the $\boldsymbol{B}$ matrix to be close to zero, because the translog reduces to a Cobb-Douglas when global concavity is imposed. A consequence of the imposition of global curvature is that it rules out most possibilities of a complementarity relationship among the inputs. Given the trade-off between theoretical regularity and functional flexibility, we thus argue in favor of an intermediate case — that of pointwise regularity (regularity at all observed data points) in that it can always guarantee the satisfaction of theoretical regularity by using the constrained optimization approach or the Bayesian approach.

References

[1] Barnett, W.A. "Tastes and Technology: Curvature is not Sufficient for Regularity." *Journal of Econometrics* 108 (2002), 199-202.

[2] Barnett, W.A. and Meenakshi Pasupathy. "Regularity of the Generalized Quadratic Production Model: A Counterexample." *Econometric Reviews* 22 (2003), 135-154.

[3] Barnett, W.A., J. Geweke, and M. Wolfe. "Semi-nonparametric Bayesian Estimation of the Asymptotically Ideal Production Model." *Journal of Econometrics* 49 (1991), 5-50.

[4] Barten, A.P. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Economic Review* 1 (1969), 7-73.

[5] Berndt, E.R. and D.O. Wood. "Technology and the Derived Demand for Energy." *Review of Economic and Statistics* 57 (1975), 259-268.

[6] Chib, S. and E. Greenberg. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49 (1995), 327-335.

[7] Chib, S. and E. Greenberg. "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory* 12 (1996), 409–431.

[8] Christensen, L.R., D.W. Jorgenson, and L.J. Lau. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65 (1975), 367-383.

[9] Diewert, W.E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." *Journal of Political Economy* 79 (1971), 481-507.

[10] Diewert, W.E. "The Duality Approach to Microeconomic Theory." In Kenneth J. Arrow and Michael D. Intriligator (eds). *Handbook of Mathematical Economics* Vol. 2, Amsterdam: North Holland (1982), pp. 535–599.

[11] Diewert, W.E. and T.J. Wales. "Flexible Functional Forms and Global Curvature Conditions." *Econometrica* 55 (1987), 43-68.

[12] Feng, G. and A. Serletis. "Productivity Trends in U.S. Manufacturing: Evidence from the NQ and AIM Cost Functions." *Journal of Econometrics* 142 (2008), 281-311.

[13] Gallant, A.R. "Unbiased Determination of Production Technology." *Journal of Econometrics* 20 (1982), 285-323.

[14] Gallant, A.R. and G. Golub. "Imposing Curvature Restrictions on Flexible Functional Forms." *Journal of Econometrics* 26 (1984), 295-321.

[15] Geman, S. and D. Geman. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1984), 609-628.

[16] Griffiths, W.E., C.J. O'Donnell, and A. Tan Cruz. "Imposing Regularity Conditions on a System of Cost and Cost-Share Equations: A Bayesian Approach." *Australian Journal of Agricultural and Resource Economics* 44 (2000), 107-127.

[17] Hastings, W.K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrica* 57 (1970), 97-109.

[18] Koop, G., J. Osiewalski, and M. Steel. "Bayesian Efficiency Analysis through Individual Effects: Hospital Cost Frontiers." *Journal of Econometrics* 76 (1997), 77-105.

[19] Lau, L.J. "Testing and Imposing Monotonicity, Convexity, and Quasiconcavity," in M. Fuss and D. McFadden (eds.), *Production Economics, A Dual Approach to Theory and Applications*, Vol(1978), Amsterdam: North-Holland, 409-453.

[20] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. "Equations of State Calculations by Fast Computing Machines" *Journal of Chemical Physics* 21 (1953), 1087-1092.

[21] Morey, E.R. "An Introduction to Checking, Testing, and Imposing Curvature Properties: The True Function and the Estimated Function." *Canadian Journal of Economics* 19 (1986), 207-235.

[22] Ryan, D.L. and T.J. Wales. "Imposing Local Concavity in the Translog and Generalized Leontief Cost Functions." *Economics Letters* 67 (2000), 253-260.

[23] Serletis, A. and A. Shahmoradi. "Semi-Nonparametric Estimates of the Demand for Money in the United States." *Macroeconomic Dynamics* 9 (2005), 542-559.

[24] Serletis, A. and A. Shahmoradi. "Flexible Functional Forms, Curvature Conditions, and the Demand for Assets." *Macroeconomic Dynamics* 11 (2007), 455-486.

[25] Sickles, R.C. "A Nonlinear Multivariate Error Components Analysis of Technology and Specific Factor Productivity Growth with an Application to the U.S. Airlines." *Journal of Econometrics* 27 (1985), 61-78.

[26] Terrell, D. "Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms." *Journal of Applied Econometrics* 11 (1996), 179-194.

[27] Tierney, L. "Markov Chains for Exploring Posterior Distributions (with discussion)" *Annals of Statistics* 22 (1994), 1701-1762.

[28] Wiley, D.E., W.H. Schmidt, and W.J. Bramble. "Studies of a Class of Covariance Structure Models." *Journal of the American Statistical Association* 68 (1973), 317-323.

[29] Zellner, A. "The Bayesian Method of Moments (BMOM): Theory and Applications." In T. Fomby and R. Hill (eds.), *Advances in Econometrics*, Vol. 12 (1997), pp. 85-105. Greenwich, CT: JAI Press.

Table 1

Parameter Estimates From The Unconstrained Model

Parameter	Estimation method: LSQ Parameter Estimate	LSQ Standard Error	Optimization Parameter Estimate	Bayesian Parameter Estimate	Bayesian Standard Deviation
β_0	-0.051	(0.014)	-0.051	-0.049	(0.018)
β_1	0.164	(0.001)	0.164	0.164	(0.002)
β_2	0.458	(0.003)	0.458	0.458	(0.004)
β_3	0.027	(0.001)	0.027	0.027	(0.001)
β_{11}	0.106	(0.004)	0.106	0.106	(0.004)
β_{12}	-0.055	(0.007)	-0.055	-0.055	(0.008)
β_{13}	-0.010	(0.002)	-0.010	-0.010	(0.002)
β_{22}	0.248	(0.019)	0.247	0.244	(0.024)
β_{23}	-0.018	(0.004)	-0.018	-0.018	(0.006)
β_{33}	0.017	(0.002)	0.017	0.017	(0.002)
β_{1t}	0.002	(0.000)	0.002	0.002	(0.000)
β_{2t}	-0.007	(0.000)	-0.007	-0.007	(0.001)
β_{3t}	0.000	(0.000)	0.000	0.000	(0.000)
β_t	-0.003	(0.006)	-0.003	-0.003	(0.002)
β_{tt}	-0.000	(0.000)	-0.000	-0.000	(0.000)
Log likelihood value	713.484		713.484	713.451	
Positivity violations	0		0	0	
Monotonicity violations	0		0	0	
Curvature violations	49		49	49	

Notes: Sample period, annual data 1953-2001 ($T = 49$). Standard errors are in parentheses.

Table 2

Parameter Estimates from Constrained LSQ

Parameter	Curvature imposed (locally) at $t = 28$	Curvature imposed globally
β_0	1.327 (0.007)	−0.002 (0.013)
β_1	0.260 (0.002)	0.170 (0.006)
β_2	0.259 (0.002)	0.477 (0.007)
β_3	0.228 (0.004)	0.013 (0.002)
β_{11}	0.169 (0.001)	−0.000 (0.000)
β_{12}	−0.071 (0.005)	0.000 (0.025)
β_{13}	−0.033 (0.006)	0.000 (0.110)
β_{22}	0.178 (0.002)	−0.000 (0.249)
β_{23}	−0.057 (0.004)	−0.000 (0.612)
β_{33}	0.091 (0.014)	−0.024 (0.745)
β_{1t}	0.003 (0.000)	0.000 (0.000)
β_{2t}	−0.005 (0.000)	−0.002 (0.002)
β_{3t}	0.001 (0.000)	0.001 (0.001)
β_t	−0.007 (0.000)	−0.008 (0.008)
β_{tt}	−0.000 (0.000)	−0.000 (0.000)
Log likelihood value	627.518	561.627
Positivity violations	0	0
Monotonicity violations	0	0
Curvature violations	6	0

Notes: Sample period, annual data 1953-2001 ($T = 49$).
Standard errors are in parentheses.

Table 3

Parameter Estimates from Constrained Optimization

Parameter	Curvature imposed at t = 16-20	Curvature imposed at every data point	Curvature imposed globally
β_0	-0.042	-0.042	-0.004
β_1	0.163	0.163	0.170
β_2	0.463	0.463	0.478
β_3	0.025	0.025	0.020
β_{11}	-0.004	-0.004	-0.008
β_{12}	0.105	0.105	0.000
β_{13}	-0.055	-0.055	0.000
β_{22}	-0.012	-0.012	0.000
β_{23}	0.211	0.211	0.000
β_{33}	-0.009	-0.009	-0.000
β_{1t}	0.013	0.013	0.000
β_{2t}	0.002	0.002	0.000
β_{3t}	-0.007	-0.007	-0.002
β_t	0.000	0.000	0.000
β_{tt}	-0.000	-0.000	-0.000
Log likelihood value	706.769	706.769	588.589
Positivity violations	0	0	0
Monotonicity violations	0	0	0
Curvature violations	0	0	0

Notes: Sample period, annual data 1953-2001 ($T = 49$). Standard errors are in parentheses.

Table 4

Parameter Estimates And 90% Posterior Credible Intervals
Under Constrained Bayesian Estimation

	Curvature imposed					
	at $t = 17$		at every data point		globally	
Parameter	Estimate	90% Posterior Credible Interval	Estimate	90% Posterior Credible Interval	Estimate	90% Posterior Credible Interval
β_0	-0.038	$(-0.066, -0.012)$	-0.039	$(-0.073, -0.013)$	-0.005	$(-0.028, 0.021)$
β_1	0.163	$(0.161, 0.165)$	0.163	$(0.161, 0.166)$	0.170	$(0.161, 0.180)$
β_2	0.463	$(0.457, 0.470)$	0.464	$(0.458, 0.470)$	0.479	$(0.469, 0.488)$
β_3	0.025	$(0.0240, 0.027)$	0.025	$(0.024, 0.026)$	0.019	$(0.017, 0.022)$
β_{11}	-0.004	$(-0.007, -0.002)$	-0.004	$(-0.007, -0.001)$	-0.008	$(-0.010, -0.007)$
β_{12}	0.106	$(0.099, 0.113)$	0.105	$(0.098, 0.111)$	-0.005	$(-0.012, -0.001)$
β_{13}	-0.056	$(-0.068, -0.044)$	-0.053	$(-0.064, -0.040)$	0.003	$(-0.004, 0.014)$
β_{22}	-0.012	$(-0.015, -0.009)$	-0.011	$(-0.015, -0.008)$	0.000	$(-0.002, 0.003)$
β_{23}	0.208	$(0.183, 0.230)$	0.201	$(0.175, 0.221)$	-0.016	$(-0.040, -0.002)$
β_{33}	-0.001	$(-0.019, -0.003)$	-0.010	$(-0.018, -0.003)$	-0.001	$(-0.006, 0.003)$
β_{1t}	0.013	$(0.009, 0.015)$	0.012	$(0.009, 0.015)$	-0.002	$(-0.004, -0.000)$
β_{2t}	0.002	$(0.002, 0.003)$	0.002	$(0.002, 0.003)$	0.000	$(-0.000, 0.000)$
β_{3t}	-0.007	$(-0.007, -0.006)$	-0.006	$(-0.007, -0.006)$	-0.002	$(-0.002, -0.001)$
β_t	0.000	$(-0.000, 0.000)$	0.000	$(-0.000, 0.000)$	0.000	$(0.000, 0.001)$
β_{tt}	-0.000	$(-0.000, 0.000)$	-0.000	$(-0.000, 0.000)$	-0.000	$(-0.000, -0.000)$
Log likelihood value	706.246		705.055		583.068	
Positivity violations	0		0		0	
Monotonicity violations	0		0		0	
Curvature violations	0		0		0	

Notes: Sample period, annual data 1953-2001 ($T = 49$).

Table 5

Price Elasticities at $t = 1$ Under The Different Methods of Estimation

Price Elasticities	LSQ Unrestricted	LSQ Curvature imposed globally	Optimization Unrestricted	Optimization Curvature imposed at $t = 16$-20	Optimization Curvature imposed at every point	Optimization Curvature imposed globally	Bayesian Unrestricted	Bayesian Curvature imposed at $t = 17$	Bayesian Curvature imposed at every point	Bayesian Curvature imposed globally
η_{11}	−0.198	−0.830	−0.198	−0.197	−0.197	−0.829	−0.197	−0.199	−0.200	−0.857
η_{12}	0.117	0.476	0.117	0.125	0.125	0.476	0.117	0.120	0.135	0.494
η_{13}	−0.031	0.014	−0.031	−0.046	−0.046	0.020	−0.030	−0.045	−0.039	0.021
η_{14}	0.112	0.340	0.112	0.118	0.118	0.334	0.110	0.124	0.105	0.342
η_{21}	0.043	0.170	0.043	0.045	0.045	0.171	0.043	0.044	0.049	0.177
η_{22}	−0.002	−0.525	−0.002	−0.080	−0.080	−0.524	−0.009	−0.087	−0.107	−0.558
η_{23}	−0.012	0.013	−0.012	0.006	0.006	0.020	−0.013	0.002	0.001	0.017
η_{24}	−0.029	0.341	−0.030	0.030	0.030	0.334	−0.021	0.041	0.058	0.363
η_{31}	−0.192	0.173	−0.192	−0.294	−0.294	0.171	−0.185	−0.296	−0.253	0.177
η_{32}	−0.195	0.455	−0.195	0.105	0.105	0.476	−0.217	0.037	0.012	0.411
η_{33}	−0.342	−2.673	−0.342	−0.456	−0.456	−0.979	−0.347	−0.493	−0.494	−1.056
η_{34}	0.728	2.045	0.728	0.645	0.645	−0.667	0.749	0.743	0.716	−0.543
η_{41}	0.052	0.170	0.052	0.056	0.056	0.171	0.051	0.058	0.049	0.176
η_{42}	−0.037	0.477	−0.038	0.038	0.038	0.476	−0.026	0.054	0.075	0.523
η_{43}	0.055	0.084	0.055	0.047	0.047	0.020	0.057	0.054	0.053	0.028
η_{44}	−0.070	−0.731	−0.070	−0.140	−0.140	−0.666	−0.082	−0.166	−0.178	−0.727

Notes: Sample period, annual data 1953-2001 ($T = 49$).

Table 6

Price Elasticities at $t = 49$ Under The Different Methods of Estimation

Price Elasticities	LSQ Unrestricted	LSQ Curvature imposed globally	Optimization Unrestricted	Optimization Curvature imposed at $t = 16$-20	Optimization Curvature imposed at every point	Optimization Curvature imposed globally	Bayesian estimation Unrestricted	Bayesian estimation Curvature imposed at $t = 17$	Bayesian estimation Curvature imposed at every point	Bayesian estimation Curvature imposed globally
η_{11}	−0.214	−0.821	−0.214	−0.215	−0.215	−0.821	−0.213	−0.216	−0.218	−0.848
η_{12}	0.069	0.383	0.069	0.071	0.071	0.382	0.069	0.066	0.080	0.398
η_{13}	−0.021	0.033	−0.021	−0.033	−0.033	0.034	−0.020	−0.033	−0.027	0.035
η_{14}	0.166	0.405	0.166	0.177	0.177	0.404	0.164	0.182	0.165	0.414
η_{21}	0.030	0.179	0.030	0.031	0.031	0.179	0.030	0.029	0.035	0.187
η_{22}	0.022	−0.618	0.022	−0.068	−0.068	−0.618	0.014	−0.076	−0.099	−0.661
η_{23}	−0.009	0.032	−0.009	0.012	0.012	0.034	−0.011	0.008	0.006	0.031
η_{24}	−0.043	0.406	−0.043	0.025	0.025	0.404	−0.033	0.039	0.058	0.444
η_{31}	−0.101	0.180	−0.102	−0.160	−0.160	0.179	−0.097	−0.160	−0.130	0.183
η_{32}	−0.104	0.374	−0.103	0.137	0.137	0.382	−0.121	0.089	0.068	0.343
η_{33}	−0.481	−1.689	−0.481	−0.591	−0.591	−0.965	−0.484	−0.619	−0.617	−1.010
η_{34}	0.687	1.135	0.686	0.614	0.614	−0.596	0.703	0.683	0.664	−0.524
η_{41}	0.071	0.179	0.071	0.076	0.076	0.179	0.070	0.078	0.071	0.184
η_{42}	−0.042	0.383	−0.042	0.024	0.024	0.382	−0.032	0.038	0.056	0.419
η_{43}	0.060	0.092	0.060	0.054	0.054	0.034	0.062	0.061	0.060	0.041
η_{44}	−0.090	−0.654	−0.090	−0.154	−0.154	−0.596	−0.100	−0.176	−0.186	−0.644

Notes: Sample period, annual data 1953-2001 ($T = 49$).

Figure 1. Estimates of the posterior pdf for η_{22}