Productivity Trends in U.S. Manufacturing: Evidence from the NQ and AIM Cost Functions*

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

Forthcoming in: Journal of Econometrics
August 21, 2007

Abstract
In this paper, we take the econometric approach to productivity measurement in United States manufacturing, using KLEM data over the period from 1953 to 2001. We are also interested in technical change bias, price elasticities, and elasticities of substitution in the U.S. manufacturing industry. We present an empirical comparison and evaluation of the effectiveness of four well-known flexible cost functions — the locally flexible generalized Leontief [see Diewert (1971)], translog [see Christensen et al. (1975)], and normalized quadratic [see Diewert and Wales (1987)] — and the globally flexible Asymptotically Ideal Model [see Barnett et al. (1991)], the latter modified to introduce technical change by means of Thomsen’s (2000) factor-augmenting efficiency index approach.

JEL classification: C22, F33.

Keywords: Flexible functional forms; Technical change; Generalized Leontief (GL); Translog; Normalized quadratic (NQ); Asymptotically Ideal Model (AIM).

*This paper builds on material from Chapter 2 of Guohua Feng’s Ph.D. dissertation at the University of Calgary. We would like to thank Clint Cummins and two referees for comments that greatly improved the paper. We also thank the members of Guohua’s dissertation committee: John Boyce, Francisco Gonzalez, Daniel Gordon, Lasheng Yuan, and Alexander David. Serletis gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC).

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

The analysis and measurement of productivity performance has attracted a great deal of attention ever since Solow (1957) decomposed the growth in output into the growth of inputs and a residual-based productivity term. Most of the literature follows the innovative works by Jorgenson and Griliches (1967), Diewert (1976), Berndt and Khaled (1979), Diewert and Wales (1987), and Fare et al. (1994) in investigating total factor productivity. This literature is interesting not only because of the critical importance of productivity growth to living standards in actual economies, but also because of the related interesting topics analyzed together with productivity growth, such as factor substitutability, convergence of macroeconomic structure and performance, and economic growth.

There are four different approaches to total factor productivity (TFP) measurement — growth accounting, the index number approach, the distance function approach, and the econometric approach. In this paper, we briefly review each of these methods, and then take the econometric approach to productivity measurement in the United States. In doing so, we use manufacturing KLEM (capital, labor, energy, and intermediate materials) data, over the period from 1953 to 2001. We are also interested in technical change bias, price elasticities, and elasticities of substitution in the U.S. manufacturing industry. We also present an empirical comparison and evaluation of the effectiveness of four well-known flexible cost functions — namely, the locally flexible generalized Leontief [see Diewert (1971)], translog [see Christensen et al. (1975)], and normalized quadratic [see Diewert and Wales (1987)], and one globally flexible cost function, the Asymptotically Ideal Model [see Barnett et al. (1991)]. In this literature, there is no a priori view as to which flexible functional forms are appropriate, once they satisfy the regularity conditions of neoclassical microeconomic theory — positivity, monotonicity, and curvature.

We pay explicit attention to all three theoretical regularity conditions and argue that much of the older literature on total factor productivity measurement ignores economic regularity. We argue that unless economic regularity is attained by luck, flexible functional forms should always be estimated subject to regularity, as suggested by Barnett (2002) and Barnett and Pasupathy (2003). In fact, we follow Ryan and Wales (1998), Moschini (1999), Gallant and Golub (1984), and Serletis and Shahmoradi (2005, 2007) and treat the curvature property as a maintained hypothesis and build it into the models being estimated. We also address econometric regularity issues and highlight the challenge inherent with achieving both economic and econometric regularity.

We also extend the AIM model and introduce technical change in the AIM cost function by means of Thomsen’s (2000) factor-augmenting efficiency index approach. The main advantage of this approach, unlike the generic time trend models of technical change, is that one can measure input specific productivity, changes in input productivity, as well as the contribution of each input to overall productivity. Our empirical results show that the AIM cost function with technical change introduced through the factor-augmenting efficiency in-
dex approach performs better than traditional locally flexible function forms and gives more accurate estimates of total factor productivity.

The rest of the paper is organized as follows. Section 2 provides a brief review of the different approaches to total factor productivity measurement — growth accounting, the index number approach, the distance function approach, and the econometric approach. In Section 3 we follow the econometric approach and discuss in detail the four cost functions that we use as well as the relevant procedures for imposing concavity on each of these functions. In Section 4 we deal with data and econometric issues while in Section 5 we estimate the models, report on theoretical regularity violations, and report estimates of total factor productivity based on the best-performing model(s). The final section concludes the paper.

2 Productivity Measurement

As already noted, there are four different approaches to total factor productivity measurement — growth accounting, the index number approach, the distance function approach, and the econometric approach. In what follows, we briefly discuss each of these approaches.

2.1 Growth Accounting

Growth accounting was suggested by Solow (1957) as a method of estimating the growth of total factor productivity. Growth accounting calculation of total factor productivity requires the specification of a neoclassical production function. Consider, for example, the Cobb-Douglas production function,

$$Y = AK^\alpha L^{1-\alpha} \quad \alpha \in (0, 1),$$

where $Y$ is (real) output, $K$ is capital, $L$ is labor, and $\alpha$ is the share of capital in output. $A$ is a measure of the current level of technology, more commonly referred to as multi-factor growth productivity or total factor productivity (TFP) — if, for example, $A$ increases by 1% and if the inputs ($K$ and $L$) are unchanged, then output increases by 1%.

As noted by Carlaw and Lipsey (2003), total factor productivity can be calculated either as a geometric index in levels,

$$TFP = \frac{Y}{K^\alpha L^{1-\alpha}} = A,$$

or as an arithmetic index in rates of change,

$$\frac{\Delta A}{A} = \frac{\Delta Y}{Y} - \alpha \frac{\Delta K}{K} + (1 - \alpha) \frac{\Delta L}{L} = \Delta TFP.$$

Equation (2) is the key equation in growth accounting. It defines the growth of total factor productivity, $\Delta A/A$, as the growth in output that cannot be accounted for by growth in
capital and labor. $\Delta A/A$ is called the Solow residual, after Robert Solow who suggested this method of estimating the growth of total factor productivity. It is also known as the rate of technical progress.

### 2.2 The Index Number Approach

The index number approach is an extension of (and complement to) growth accounting. It involves dividing a (real) output quantity index, $Y$, by an input quantity index, $I$, to obtain a measure of total factor productivity, $A$, as follows

$$A = \frac{Y}{I}.$$  

The index number approach is widely used by the majority of statistical agencies that regularly produce productivity statistics. However, one critical issue regarding this approach is the selection of the appropriate indexes. In fact, statistical indexes are mainly characterized by their statistical properties. These properties were examined in great detail by Fisher (1922) and serve as tests in assessing the quality of a particular statistical index. They have been named, after Fisher, as ‘Fisher’s system of tests’ — see Eichhorn (1976) for a detailed analysis as well as a comprehensive bibliography of Fisher’s ‘test’ or ‘axiomatic’ approach to index numbers.

The index that Fisher (1922) found to be the best, in the sense of possessing the largest number of desirable statistical properties, has now become known as the ‘Fisher ideal’ index. Another index found to possess a very large number of such properties is the discrete time approximation to the continuous Divisia index, usually called the Törnqvist index or just the Divisia index (in discrete time). In fact, the primary advantage of the Fisher ideal index over the Divisia index is that the Fisher ideal index satisfies Fisher’s ‘factor reversal test’ — which requires that the product of the price and quantity indexes for an aggregated good should equal actual expenditures on the component goods — while the discrete time approximation of the Divisia index fails that test. However, the magnitude of the error is very small — third order in the changes.

The index number approach does not require an aggregate production function, although the economic approach to statistical index numbers, pioneered by Diewert (1976), could be used for selecting the appropriate index — see also Diewert and Lawrence (1999) and Diewert and Nakamura (2003) for a detailed discussion. In particular, Diewert (1976) provided the link between aggregation theory and statistical index number theory by attaching economic properties to statistical indexes. These properties are defined in terms of the statistical indexes’ ability to approximate a particular functional form for the unknown underlying aggregator function. For example, Diewert (1976) showed that the Divisia index is ‘exact’ for the linearly homogeneous translog and is, therefore, ‘superlative’ (since the translog is a flexible functional form).
2.3 The Distance Function Approach

The distance function approach to measuring total factor productivity seeks to separate total factor productivity in two components: changes resulting from a movement towards the production frontier (technical efficiency) and shifts in the frontier (technical change). The distance function was first introduced separately by Shephard (1953) in the context of production analysis and by Malmquist (1953) in the context of consumption analysis. But it was introduced as a theoretical productivity index by Caves et al. (1982), and then popularized as an empirical productivity index by Färe et al. (1994).

Using technology in period $t$ as the reference technology (which exhibits constant returns to scale), the output-based Malmquist productivity index is written as

$$m_t^o (y_t, y_{t+1}, x_t, x_{t+1}) = \frac{d_o^r (y_{t+1}, x_{t+1})}{d_o^r (y_t, x_t)},$$

where $d_o^r (y_t, x_t)$ is the output distance function; that is, the reciprocal of the maximum proportional expansion of the output vector $y_t$, given inputs $x_t$. Alternatively, the Malmquist index can be defined in terms of technology in time $t + 1$. Färe et al. (1994) extend this approach by defining the Malmquist total factor productivity index as the geometric mean of these two indexes

$$m_t^t (y_t, y_{t+1}, x_t, x_{t+1}) = \left[ \frac{d_o^r (y_{t+1}, x_{t+1})}{d_o^r (y_t, x_t)} \times \frac{d_o^{r+1} (y_{t+1}, x_{t+1})}{d_o^{r+1} (y_t, x_t)} \right]^{1/2},$$

which can be equivalently written as

$$m_t^t (y_t, y_{t+1}, x_t, x_{t+1}) = \frac{d_o^{r+1} (y_{t+1}, x_{t+1})}{d_o^r (y_t, x_t)} \times \left[ \frac{d_o^r (y_{t+1}, x_{t+1})}{d_o^{r+1} (y_{t+1}, x_{t+1})} \times \frac{d_o^{r+1} (y_t, x_t)}{d_o^r (y_t, x_t)} \right]^{1/2},$$

where the term outside the brackets on the right-hand side measures the change in relative efficiency between years $t$ and $t + 1$ and the geometric mean of the two ratios inside the brackets measures the shift in technology between the two periods evaluated at $x_t$ and $x_{t+1}$. It is to be noted that the Malmquist total factor productivity index can also be measured on the best practice technologies when variable returns to scale are taken into account.

1There is also a closely related literature on firm efficiency, using stochastic production (or cost/profit) frontiers. Like the Malmquist productivity index discussed below, the parametric stochastic frontier approach does not assume that firms are operating at their efficient level, and thus enables one to decompose the combined productivity changes into efficiency movements (efficiency change) and frontier shift components (technological change), among other components. A two-component composite error term is usually assumed in the estimation of the parametric model with one capturing firm inefficiency and the other capturing statistical noise. For an excellent review of this literature, see Kumbhakar and Lovell (2003).
The Färe et al. (1994) distance function based productivity index has several advantages. It does not require a specific functional form, it does not require information on prices, and it can be implemented in a multiple-output setting with many inputs (with no separability assumptions being required). Most importantly, it does not assume that firms are operating at their efficient level, and thus enables one to decompose the combined productivity changes into efficiency movements and frontier shift components. However, as Carlaw and Lipsey (2003, pp. 464) put it, “in order to implement this technique, one must know everything about the state of technology at every point in time and at every level of aggregation that TFP is calculated. Unfortunately, this is not possible given the data available.” Moreover, implicit in the distance function approach to measuring total factor productivity is the assumption that all units (firms, industries, or countries) being compared have the same production function, when in fact evidence suggests that even firms within the same industry do not have identical production functions.

2.4 The Econometric Approach

The econometric approach to productivity measurement involves estimating the parameters of an aggregator function — cost, profit, or production function. Productivity growth can then be expressed in terms of the estimated parameters.

Technical change (or productivity growth) is usually defined in the primal setup (production function), as any shift in the production frontier. In particular, assuming a production function

\[ y = f(x, t), \]

where \( y \) is output, \( f \) is a continuous twice differentiable nondecreasing and quasiconcave function of a vector of inputs \( x \geq 0 \), and \( t \) denotes a technology index, then technical change is defined as

\[ \frac{\partial f(x, t)}{\partial t}. \]

Technical change can also be defined 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 (3) is completely described by the dual cost function

\[ C = C(p, y, t) = yc(p, t), \]

with the second equality assuming constant returns to scale. In equation (4), \( C \) is a non-decreasing, linearly homogeneous and concave function of prices, \( p > 0 \), and \( c \) is the corresponding unit cost function — for an excellent review of duality theory, see Diewert (1982).

To obtain equations that are amenable to estimation, we apply Shephard’s lemma to equation (4) to get

\[ x_i = \frac{\partial C(p, y, t)}{\partial p_i}, \quad i = 1, \ldots, n, \]
or a more convenient equation for estimation purposes, by dividing through by \( y \),

\[
\frac{x_i}{y} = \frac{1}{y} \frac{\partial C(p, y, t)}{\partial p_i}, \quad i = 1, \ldots, n.
\]

Using the envelope theorem,

\[
\frac{\partial C(p, y, t)}{\partial t} = -\frac{\partial C(p, y, t)}{\partial y} \frac{\partial f(x, t)}{\partial t},
\]

the rate of technical change can be measured from the cost function as follows

\[
TFP = \frac{\partial \ln f(x, t)}{\partial t} = \frac{1}{y} \frac{\partial f(x, t)}{\partial t}
\]

\[
= -\frac{\partial C(p, y, t)}{y \partial C(p, y, t) / \partial y} = -\frac{\partial \ln C(p, y, t) / \partial t}{\partial \ln C(p, y, t) / \partial \ln y}
\]

\[
= -\epsilon_{ct}\epsilon_{cy}^{-1}, \tag{6}
\]

where \( \epsilon_{ct} = \partial \ln C(p, y, t) / \partial t \) and \( \epsilon_{cy} = \partial \ln C(p, y, t) / \partial \ln y \). According to equation (6), total factor productivity is the product of the dual rate of cost diminution (\( \epsilon_{ct} \)) and the dual rate of returns to scale (\( \epsilon_{cy}^{-1} \)). Hence, under constant returns to scale (where \( \epsilon_{cy} \) is equal to unity), total factor productivity is the negative of the dual rate of cost diminution, meaning that a 1\% upward shift in the production function is equal to a 1\% decrease in the cost of production.

By taking the derivative of each estimated factor demand equation with respect to time and dividing by the estimated demands, we can also obtain a measure of the effect of technical change on each input (denoted by \( \tau_i \) below) — see, for example, Diewert and Wales (1992) and Kohli (1994) — as follows,

\[
\tau_i = \frac{\partial \ln x_i(p, y, t)}{\partial t}. \tag{7}
\]

If \( \tau_i > 0 \) (\( \tau_i < 0 \)), then technical change is input \( i \) augmenting (reducing), meaning that more (less) of the input is required due to the passing of time. In fact, total factor productivity is a weighted average of \( \tau_i \)'s. Following Kohli (1994), we define these \( \tau_i \)'s to be technical change biases for each input. In particular, if

\[
TFP = -\tau_i, \tag{8}
\]

for input \( i \), then technical change is said to be completely unbiased (neutral), in the sense that all goods are affected to the same degree. This corresponds to a ‘homothetic shift’ of
the isoquants leaving the marginal rate of substitution between any two inputs (measured along a ray through the origin) unaffected by technical change. If, however, (8) does not hold, then technical change is said to be biased. This corresponds to a ‘non-homothetic shift’ of the isoquants, meaning that the marginal rate of substitution between any two inputs is affected by technical change.

Factor substitution is calculated, using both Allen and Morishima elasticities of substitution. The Allen-Uzawa elasticity of substitution between inputs $i$ and $j$ is given by

$$\sigma^a_{ij}(p, y, t) = \frac{C(p, y, t) C_{ij}(p, y, t)}{C_i(p, y, t) C_j(p, y, t)}, \quad (9)$$

where the $i, j$ subscripts refer to the first and second partial derivatives of $C(p, y, t)$ with respect to input prices $p_i$ and $p_j$. The Morishima elasticity of substitution between inputs $i$ and $j$ is given by

$$\sigma^m_{ij}(p, y, t) = \frac{p_j C_{ij}(p, y, t)}{C_i(p, y, t)} - \frac{p_j C_{jj}(p, y, t)}{C_j(p, y, t)}. \quad (10)$$

If $\sigma^a_{ij} > 0$ (that is, if increasing the $j^{th}$ price increases the optimal quantity of input $i$), we say that inputs $i$ and $j$ are Allen-Uzawa (net) substitutes. If $\sigma^a_{ij} < 0$, they are Allen-Uzawa (net) complements. Similarly, if $\sigma^m_{ij} > 0$ (that is, if increasing the $j^{th}$ price increases the optimal quantity of input $i$ relative to the optimal quantity of input $j$), we say that input $j$ is a Morishima (net) substitute for input $i$. If $\sigma^m_{ij} < 0$, input $j$ is a Morishima net complement to input $i$. The Allen elasticities provide immediate qualitative comparative-static information about the effect of price changes on absolute input shares, whereas the Morishima elasticities immediately yield both qualitative and quantitative information about the effect of price changes on relative input shares.

The familiar price elasticities,

$$\eta_{ij} = \frac{\partial x_i(p, y, t)}{\partial p_j} \frac{p_j}{x_i(p, y, t)}, \quad (11)$$

could also be calculated as

$$\eta_{ij} = s_j \sigma^a_{ij},$$

where $s_j$ is the cost share of input $j$ in total production costs. Notice that the price elasticities must satisfy the following condition

$$\sum_{j=1}^{n} \eta_{ij} = 0, \quad i = 1, \ldots, n.$$

By substituting different unit cost functions into (4), we can get different total cost functions. Clearly, the econometric approach overcomes the problems of the index number
approach and the distance function approach and has the flexibility to incorporate pertinent features of the market and industry structures as well as technological features that affect the productivity of firms or industries.

In what follows, we take the econometric approach to productivity measurement (in the United States) and provide a comparison between three widely used locally flexible cost functional forms — the generalized Leontief, basic translog, and normalized quadratic — and the globally flexible AIM cost function.

3 Flexible Cost Functional Forms

3.1 The Generalized Leontief Cost Function

By substituting the GL unit cost function [see Diewert (1971)] into (4), we get the GL specification

\[ C(p, y, t) = y \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} p_i^{1/2} p_j^{1/2} + \sum_{i=1}^{n} \beta_{it} p_i t \right), \]  

(12)

where \( \beta_{ij} = \beta_{ji} \). Using Shephard’s lemma (5), and dividing through by \( y \), yields optimal input-output demand equations, as follows

\[
\frac{x_i}{y} = \sum_{j=1}^{n} \beta_{ij} p_j^{1/2} p_i^{-1/2} + \beta_{it} t, \quad i = 1, \cdots, n. \tag{13}
\]

Notice that all the parameters of the GL cost function (12) can be obtained by estimating only (13). It is to be noted that when \( i = j \) in (13), \( p_j^{1/2} p_i^{-1/2} = 1 \) and so \( \beta_{ii} \) is a constant term in the \( i \)th input-output equation. When \( \beta_{ij} = 0 \) for all \( i, j, i \neq j \), then input-output demand equations are independent of relative prices and the cross-price elasticities are zero.

Caves and Christensen (1980) have shown that the GL has satisfactory local properties when technology is nearly homothetic and substitution is low. However, when technology is not homothetic and substitution increases, they show that the GL has a rather small regularity region.

Concavity of the cost function (12) requires that the Hessian matrix is negative semidefinite. We can therefore impose local concavity (that is, at the reference point) by evaluating the Hessian terms of (12) at the reference point, where all prices and output are unity, as follows

\[
H_{ij} = -\delta_{ij} \left( \sum_{j=1, j \neq i}^{n} \frac{\beta_{ij}}{2} \right) + (1 - \delta_{ij}) \beta_{ij}/2, \tag{14}
\]
where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. By replacing $H$ by $-1/2 \, KK'$, where $K$ is an $n \times n$ lower triangular matrix and $K'$ its transpose, the above can be written as

$$
-\frac{1}{2} (KK')_{ij} = -\delta_{ij} \left( \sum_{j=1,j\neq i}^{n} \beta_{ij}/2 \right) + (1 - \delta_{ij}) \beta_{ij}/2. \tag{15}
$$

There are two things that should be noted here. First, the $\beta_{ii}$ ($i = 1, \cdots, n$) do not appear in (15), thus leaving $\beta_{ii}$ ($i = 1, \cdots, n$) unrestricted. Second, the fact that the elements in the same row of $H$ add to zero, that is

$$
\sum_{j=1}^{n} H_{ij} = -\left( \sum_{j=1,j\neq i}^{n} \beta_{ij}/2 \right) + \sum_{j=1,j\neq i}^{n} \beta_{ij}/2 = 0, \quad i = 1, \cdots, n,
$$

implies the following restrictions on $K$

$$
\sum_{i=1}^{n} k_{ij} = 0, \quad j = 1, \cdots, n, \tag{16}
$$

i.e. the elements in the same column of $K$ add to zero, where the $k_{ij}$ terms are the elements of the replacement matrix $K$. (16) can be easily shown by expanding out (15); a similar technique is also used by Fox and Diewert (1999) in imposing convexity of a Normalized Quadratic profit function in prices. Obtaining the main diagonal elements of $K$, $k_{ii}$, expressed in terms of $k_{ij}$ ($i \neq j$) and then substituting them into (15), we will obtain $\beta_{ij}$ ($1 \leq i < j \leq n$) which are expressed only in terms of $k_{ij}$ ($1 \leq j < i \leq n$).

As an example, for the case of three inputs ($n = 3$), we can use the restrictions (16) and the lower triangular structure of $K$ in order to eliminate the diagonal elements of $K$, $k_{ii}$ ($i = 1, 2, 3$), as follows

$$
k_{11} = -k_{21} - k_{31};
k_{22} = -k_{32};
k_{33} = 0.
$$

Substituting the above restrictions in (15), we obtain

$$
\beta_{12} = -k_{21}k_{11} = k_{21}(k_{21} + k_{31});
\beta_{13} = -k_{31}k_{11} = k_{31}(k_{21} + k_{31});
\beta_{23} = -(k_{21}k_{31} + k_{22}k_{32}) = -k_{21}k_{31} + k_{32}^2,
$$

which guarantees concavity of the cost function at the reference point and may also induce concavity of the cost function at other data points. As already noted above, $\beta_{11}$, $\beta_{22}$, and $\beta_{33}$ in this example are unrestricted and do not have to be expressed in terms of the elements of $K$. Clearly, the flexibility of the GL is not destroyed because the $n(n-1)/2$ elements of $K$ just replace the $n(n-1)/2$ elements of $H$ in the estimation.
### 3.2 The Translog Cost Function

The translog specification, due to Christensen et al. (1975), is obtained by substituting the translog unit cost function into (4) to get

\[
\ln C(p, y, t) = \ln y + \beta_0 + \beta_t t + \sum_{i=1}^{n} \beta_i \ln p_i + \\
\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^{n} \beta_{it} \ln p_i + \frac{1}{2} \beta_{tt} t^2, \tag{17}
\]

where \( \beta_{ij} = \beta_{ji} \). Homogeneity of degree one in prices (given \( y \)) implies the following restrictions

\[
\sum_{i=1}^{n} \beta_i = 1, \quad \sum_{i=1}^{n} \beta_{ij} = \sum_{j=1}^{n} \beta_{ji} = \sum_{i=1}^{n} \beta_{it} = 0. \tag{18}
\]

Although we could estimate (17) directly, efficiency gains can be realized by estimating the optimal cost-minimizing input demand equations, transformed into cost-share equations, as follows

\[
s_i = \frac{p_i x_i}{C} = \beta_i + \sum_{j=1}^{n} \beta_{ij} \ln p_j + \beta_{it} t, \tag{19}
\]

where \( \sum_{i=1}^{n} p_i x_i = C \).

Guilkey et al. (1983) show that the translog is globally regular if and only if technology is Cobb-Douglas. In other words, the translog performs well if substitution between all factors is close to unity. They also show that the regularity properties of the translog model deteriorate rapidly when substitution diverges from unity.

The Hessian matrix of the translog cost function at the reference point, where all prices and output are set to one, will be negative semidefinite if the following matrix is negative semidefinite

\[
H_{ij} = \beta_{ij} + \beta_i \beta_j - \delta_{ij} \beta_i, \quad i, j = 1, \ldots, n, \tag{20}
\]

with \( \delta_{ij} = 1 \) if \( i = j \) and 0 otherwise. Local concavity can be imposed at the reference point as in Ryan and Wales (2000) by setting \( H = -KK' \), as follows

\[
\beta_{ij} + \beta_i \beta_j - \delta_{ij} \beta_i = (-KK')_{ij}, \quad i, j = 1, \ldots, n, \tag{21}
\]

where (as before) \( K \) is a lower triangular matrix. Noting that \( \sum_{j=1}^{n} \beta_{ij} = 0 \) and \( \sum_{j=1}^{n} \beta_j = 0 \) (see equation (18)), it can be easily shown that

\[
\sum_{j=1}^{n} H_{ij} = \sum_{j=1}^{n} (\beta_{ij} - \beta_i \delta_{ij} + \beta_i \beta_j) = 0, \tag{22}
\]
i.e. the elements in the same row of $H$ add to zero. Further, (22) implies the following restriction on the elements of $K$

$$\sum_{i=1}^{n} k_{ij} = 0, \quad j = 1, \ldots, n,$$

(23)

i.e., the elements in the same column of $K$ add to zero. Again, (23) can be shown by expanding out $H = -KK'$, where $H$ satisfies (22). Combining (21) and (23), we can replace the elements of $B = [\beta_{ij}]$ by those of $K$. It should be noted that, unlike in the case of the generalized Leontief, $\beta_{ii}$ ($i = 1, \ldots, n$) are restricted in this case.

For the case with three inputs ($n = 3$), equations (21) and (23) imply the following restrictions on the elements of $K$

$$\begin{align*}
\beta_{11} &= -k_{11}^2 + \beta_1 - \beta_2^2 = -(k_{21} + k_{31})^2 + \beta_1 - \beta_2^2; \\
\beta_{12} &= -k_{11}k_{21} - \beta_1\beta_2 = (k_{21} + k_{31})k_{21} - \beta_1\beta_2; \\
\beta_{13} &= -k_{11}k_{31} - \beta_1\beta_3 = (k_{21} + k_{31})k_{31} - \beta_1\beta_3; \\
\beta_{22} &= -(k_{21}^2 + k_{22}^2) + \beta_2 - \beta_2^2 = -k_{21}^2 - k_{32}^2 + \beta_2 - \beta_2^2; \\
\beta_{23} &= -(k_{21}k_{31} + k_{22}k_{32}) - \beta_2\beta_3 = -k_{21}k_{31} + k_{32}^2 - \beta_2\beta_3; \\
\beta_{33} &= -(k_{31}^2 + k_{32}^2 + k_{33}^2) + \beta_3 - \beta_3^2 = -(k_{31}^2 + k_{32}^2) + \beta_3 - \beta_3^2,
\end{align*}$$

which guarantee concavity of the cost function at the reference point and may also induce concavity of the cost function at other data points. Clearly, the flexibility of the translog specification is not destroyed because the $n(n-1)/2$ elements of $K$ just replace the $n(n-1)/2$ elements of $B$ in the estimation.

### 3.3 The Normalized Quadratic Cost Function

The NQ model, due to Diewert and Wales (1987), can be obtained by substituting the NQ unit cost function into (4)

$$C(p, y, t) = y \left[ \sum_{i=1}^{n} \beta_i p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} p_i p_j \sum_{i=1}^{N} \alpha_i p_i + \sum_{i=1}^{n} \beta_{it} p_i t \right],$$

(24)

where we impose two restrictions on the $B \equiv [\beta_{ij}]$ matrix

$$\beta_{ij} = \beta_{ji}, \quad \text{for all } i, j;$$

(25)

$$BP^* = 0, \quad \text{for some } p^* > 0.$$

(26)

Further, the $\alpha$ vector ($\alpha > 0$) is usually predetermined.
For the NQ cost function, the unknown parameters in (24) can be estimated by using the following system of factor demands

\[
x_i = y \beta_i + \sum_{j=1}^{n} \beta_{ij} \frac{p_i}{\sum_{i=1}^{n} \alpha_i p_i} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \frac{p_i}{\sum_{i=1}^{n} \alpha_i p_i} \frac{p_j}{\sum_{j=1}^{n} \alpha_j p_j} \right) + \beta_i t.
\]

Before estimating the system in (27), we express the main diagonal elements of the \( B \) matrix, \( \beta_{ii} \), in terms of its off-diagonal elements by using equation (26) and assuming that \( p^* = 1_n \).

Thus, by estimating the input-output equations (27), we obtain estimates of \( \beta_i \), the technical change parameters \( \beta_t \), and the off-diagonal elements of the \( B \) matrix, \( \beta_{ij} (i \neq j) \). The main diagonal elements of the \( B \) matrix can be recovered from the restrictions imposed.

The Hessian matrix of the cost function (24) is obtained as follows

\[
\nabla_{p_ip_j} C (p, y, t) = \frac{\beta_{ij}}{\sum_{i=1}^{n} \alpha_i p_i} \frac{\alpha_i \left( \sum_{j=1}^{n} \beta_{ij} p_j \right)}{\left( \sum_{i=1}^{n} \alpha_i p_i \right)^2} \left( \sum_{i=1}^{n} \alpha_i p_i \right) + \frac{\alpha_i \alpha_j \left( \sum_{i=1}^{n} \sum_{j=1}^{n} p_i \beta_{ij} p_j \right)}{\left( \sum_{i=1}^{n} \alpha_i p_i \right)^3}.
\]

Using the restrictions \( \sum_{j=1}^{n} \beta_{ij} p_j^* = 0_n \) at the reference point, we have \( \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^* \beta_{ij} p_j^* = \sum_{i=1}^{n} \left( p_i^* \left( \sum_{j=1}^{n} \beta_{ij} p_j^* \right) \right) = 0 \). Thus evaluating the above equation at \((p^*, t^*)\) yields the following equation

\[
\nabla_{p_ip_j} C (p, y, t) = \frac{\beta_{ij}}{\sum_{i=1}^{n} \alpha_i p_i^*}.
\]

Multiplying both sides of (29) by \( y \) and rearranging, we get \( \nabla_{p_ip_j} C (p, y, t) = \alpha' p^{-1} B \).

Thus the negative semidefiniteness of \( \nabla_{p_ip_j} C (p, y, t) \) at the reference point requires that \( B \) is negative semidefinite. More importantly, the negative semidefiniteness of \( B \) is not only the necessary condition for \( \nabla_{p_ip_j} C (p, y, t) \) to be concave locally at the reference point as we just showed, but it is also a sufficient condition for \( \nabla_{p_ip_j} C (p, y, t) \) to be concave globally (concave at every possible and imaginable point) — see Diewert and Wales (1987) for more details.

In practice, the concavity of \( C (p, y, t) \) may not be satisfied, in the sense that the estimated \( B \) matrix may not be negative semidefinite. In this case, to ensure global concavity (concavity at all possible prices) of the NQ cost function, we follow Diewert and Wales (1987) and impose

\[
B = -K K',
\]

where \( K \) is a lower triangular matrix which satisfies

\[
K' p^* = 0_n.
\]
Note that (31) and the lower triangular structure of $K$ imply

$$\sum_{i=1}^{n} k_{ij} = 0, \quad j = 1, \ldots, n. \quad (32)$$

As an example, for the case of three inputs (30) and (32) imply

$$\begin{align*}
\beta_{11} &= -k_{11}^2 = -(k_{21} + k_{31})^2; \\
\beta_{12} &= -k_{11}k_{21} = (k_{21} + k_{31})k_{21}; \\
\beta_{13} &= -k_{11}k_{31} = (k_{21} + k_{31})k_{31}; \\
\beta_{22} &= -(k_{21}^2 + k_{22}^2) = -k_{21}^2 - k_{32}^2; \\
\beta_{23} &= -(k_{21}k_{31} + k_{22}k_{32}) = -k_{21}k_{31} + k_{32}^2; \\
\beta_{33} &= -(k_{31}^2 + k_{32}^2 + k_{33}^2) = -(k_{31}^2 + k_{32}^2). 
\end{align*}$$

That is, we replace the elements of $B$ in the input-output equations (27) by the elements of $K$, thus ensuring global curvature. It should be noted that in the case of the NQ cost model, concavity is imposed globally rather than locally at the reference point as we do in the case of the GL and translog specifications. The main advantage of the NQ specification comes from its property that correct curvature conditions can be imposed globally without destroying the flexibility of the functional form.

### 3.4 The AIM Cost Function

By specifying the unit cost function in (4) as a linearly homogeneous multivariate Müntz-Szatz series expansion, we get the AIM total cost function without technical change [see Barnett et al. (1991)]

$$C = g(p, y) = y \left[ \sum_{z \in A_\kappa} a_z \prod_{j=1}^{2^\kappa} \rho_{i_j}^{2^\kappa} \right], \quad (33)$$

where $\kappa$ is the order of expansion, $a_z$ the unknown parameters, $n$ the number of production factors, and $A_\kappa = \{(i_1, i_2, \ldots, i_{2^\kappa}) : i_1, i_2, \ldots, i_{2^\kappa} \in \{1, 2, \ldots, n\}; i_1 \leq i_2, \leq \cdots \leq i_{2^\kappa}\}$. For simplicity, we call a cost function without technical change the ‘stripped-down cost function.’

Now we extend the Barnett et al. (1991) stripped-down AIM cost function to allow for technical change — a very valuable source of information about modeling technical change is Sato (1975). Instead of using the generic time trend to model technical change, as we did with the three locally flexible functional forms, we introduce technical change into the stripped-down AIM cost function using the efficiency index approach. In particular, we assume that the effects of technical change $(t)$ on the production level $y$ are purely factor-augmenting; that is, affecting each factor through a factor specific efficiency index, $e_i = e_i(t, y)$ — factor augmenting technical change was pioneered by Kohli (1981, 1982, 1991, 1993).
Thomsen (2000) shows generally that in order to obtain a total cost function with technical change and returns to scale, \( C(p, y, t) \), one can first figure out the stripped-down cost function denoted by \( C^*(p, y) \), and then divide \( p \) in \( C^*(p, y) \) by a factor specific efficiency index. Thomsen (2000) further shows that the efficiency index is capable of rendering any stripped-down cost function flexible in \( y \) and \( t \). Under the assumption of constant returns to scale we specify the efficiency index as\(^2\)

\[
\log e_i = \vartheta_i t \quad i = 1, \ldots, n,
\]

where \( \vartheta_i \) indicates (when multiplied by 100) the percentage increase in the efficiency of factor \( i \) from period \( t \) to \( t + 1 \). By dividing \( p \) in (33) by the above efficiency index, we obtain our AIM cost function with technical change

\[
C = g(p, y, t) = y \left[ \sum_{z \in A_n} a_z \prod_{j=1}^{2^\kappa} q^{2^{2^\kappa-1}} \right] = yQ(q),
\]

where \( q_{ij} \) is the efficiency-corrected price, defined as \( q_{ij} = p_{ij}/e_{ij} \), \( e \) is the efficiency index as defined by (34), and \( Q(q) \) is the corresponding unit cost function. The main advantage of the efficiency index approach is that we can easily obtain a new AIM cost function with technical change which retains all of the theoretical properties of the stripped-down AIM cost function. Another advantage of this approach is that one can measure input-specific productivity, changes in input productivity, and the contribution of each input to overall productivity, unlike the generic time trend models of technical change. We shall discuss these advantages in more detail in what follows.

Our AIM total cost function with technical change retains all the theoretical properties of the Barnett et al. (1991) stripped-down AIM cost function. First, our AIM total cost function with technical change is still globally flexible in the sense that it is capable of approximating the underlying cost function at every point in the function’s domain by increasing the order of expansion \( \kappa \). Second, it can be clearly seen from (35) that the sum of the exponents of prices in each term is still \( 2^{2^\kappa-1} = 1 \), thus satisfying the property of global linear homogeneity. Third, as with the stripped-down AIM cost function, we can impose concavity and monotonicity on the coefficients of our AIM total cost function with technical change by requiring all the coefficients to be nonnegative. In particular, with nonnegative coefficients, the function \( \left( p_{ij} e_{ij}^{-\vartheta_{ij,t}} \right)^{2^{2^\kappa-1}} \) is increasing and concave in \( p \) for any fixed \( \kappa \). Hence,

\(^2\)Returns to scale can be easily incorporated in the AIM cost function by modifying the efficiency index. Regarding the assumption of constant returns to scale, see the description of the data in Section 4.
according to Berge (1963, Theorem 1), \( \prod_{j=1}^{2^\kappa} \left( p_{ij} e_{ij}^{-\theta_{ij} t} \right)^{2^{-\kappa}} \) is increasing and quasiconcave jointly in all of its variables for any fixed \( \kappa \). However, as shown by Diewert and Wales (1993) when global concavity is imposed on this functional form in this manner, it is not flexible and complements are ruled out.

Applying Shephard’s lemma (5) to (35), and dividing through by \( y \), yields optimal input-output demand equations, as follows

\[
\frac{x_i}{y} = \frac{\partial}{\partial p_i} \left( \sum_{z \in A_k} a_z \prod_{j=1}^{2^\kappa} \left( p_{ij} e_{ij}^{-\theta_{ij} t} \right)^{2^{-\kappa}} \right). \tag{36}
\]

The system of factor demand functions produced by applying Shephard’s lemma to the \( k \)th partial sum of the cost function, \( C_k (p, y, t) \) will be called the AIM(\( k \)) factor demand system, and the resulting input-output equations will be called the AIM(\( k \)) input-output system.

In empirical applications, the approximation of the AIM cost function must be truncated at some finite value \( \kappa \) (i.e. finite partial sums). The order of approximation \( \kappa \) is usually determined empirically and stops when the elasticity estimates and the covariance matrix of the disturbances converge. By using formula (35) and (36), we now explicitly produce the first two partial sum of our expansion of the cost function in the four-factor case, AIM(1) and AIM(2) respectively.

For four goods \((n = 4)\) and \( \kappa = 1 \), the AIM(1) cost function can be written as

\[
C_{\kappa=1} (p, y, t) = y \left( \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3 + \alpha_4 q_4 + \alpha_5 q_1^{1/2} q_2^{1/2} + \alpha_6 q_3^{1/2} q_4^{1/2} + \alpha_7 q_1^{1/2} q_4^{1/2} + \alpha_8 q_2^{1/2} q_3^{1/2} + \alpha_9 q_2^{1/2} q_4^{1/2} + \alpha_{10} q_3^{1/2} q_4^{1/2} \right). \tag{37}
\]

Applying Shephard’s lemma (5) to (37) yields the factor demand equations of the AIM(1) model

\[
\begin{align*}
\frac{x_1}{y} & = \frac{1}{e^{\theta_{11} t}} \left( \alpha_1 + \frac{1}{2} \alpha_5 q_1^{1/2} q_2^{1/2} + \frac{1}{2} \alpha_6 q_1^{1/2} q_3^{1/2} + \frac{1}{2} \alpha_7 q_1^{1/2} q_4^{1/2} \right); \\
\frac{x_2}{y} & = \frac{1}{e^{\theta_{21} t}} \left( \alpha_2 + \frac{1}{2} \alpha_5 q_1^{1/2} q_2^{1/2} - q_2^{1/2} - q_1^{1/2} - q_3^{1/2} + \frac{1}{2} \alpha_8 q_2^{1/2} q_3^{1/2} + \frac{1}{2} \alpha_9 q_2^{1/2} q_4^{1/2} \right); \\
\frac{x_3}{y} & = \frac{1}{e^{\theta_{31} t}} \left( \alpha_3 + \frac{1}{2} \alpha_6 q_1^{1/2} q_3^{1/2} - q_3^{1/2} - q_1^{1/2} - q_2^{1/2} + \frac{1}{2} \alpha_8 q_2^{1/2} q_3^{1/2} + \frac{1}{2} \alpha_9 q_2^{1/2} q_4^{1/2} \right); \\
\frac{x_4}{y} & = \frac{1}{e^{\theta_{41} t}} \left( \alpha_4 + \frac{1}{2} \alpha_7 q_1^{1/2} q_4^{1/2} + \frac{1}{2} \alpha_8 q_2^{1/2} q_4^{1/2} + \frac{1}{2} \alpha_9 q_2^{1/2} q_4^{1/2} \right). \end{align*}
\]
For $n = 4$ and $k = 2$, the AIM(2) cost function can written as

$$C_{k=2}(p, y, t) = y \left( \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3 + \alpha_4 q_4 + \alpha_5 q_1^{1/2} q_2^{1/2} + \alpha_6 q_1^{1/2} q_3^{1/2} + \alpha_7 q_1^{1/2} q_4^{1/2} + \alpha_8 q_2^{1/2} q_3^{1/2} + \alpha_9 q_2^{1/2} q_4^{1/2} + \alpha_{10} q_3^{1/2} q_4^{1/2} + \alpha_{11} q_1^{3/4} q_2^{1/4} + \alpha_{12} q_1^{1/4} q_3^{3/4} + \alpha_{13} q_1^{3/4} q_4^{1/4} + \alpha_{14} q_1^{1/4} q_3^{3/4} + \alpha_{15} q_1^{3/4} q_4^{1/4} + \alpha_{16} q_1^{1/4} q_3^{3/4} + \alpha_{17} q_2^{3/4} q_3^{1/4} + \alpha_{18} q_2^{1/4} q_3^{3/4} + \alpha_{19} q_2^{3/4} q_3^{1/4} + \alpha_{20} q_2^{1/4} q_4^{3/4} + \alpha_{21} q_3^{3/4} q_4^{1/4} + \alpha_{22} q_3^{1/4} q_4^{3/4} + \alpha_{23} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{24} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{25} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{26} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{27} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{28} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{29} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{30} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{31} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{32} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{33} q_1^{1/2} q_2^{1/2} q_3^{1/4} + \alpha_{34} q_1^{1/2} q_2^{1/2} q_4^{1/2} + \alpha_{35} q_1^{1/2} q_2^{1/2} q_3^{1/4} \right). \quad (38)$$

Applying (5) to (38) yields the following system of factor demand equations for the AIM(2) model

$$x_1 = \frac{1}{y} \left( \alpha_1 + \frac{1}{2} \alpha_5 q_1^{1/2} q_2^{1/2} + \frac{1}{2} \alpha_6 q_1^{1/2} q_3^{1/2} + \frac{1}{2} \alpha_7 q_1^{1/2} q_4^{1/2} + \frac{3}{4} \alpha_{11} q_1^{1/4} q_2^{1/4} + \frac{1}{4} \alpha_{12} q_1^{1/4} q_3^{3/4} + \frac{3}{4} \alpha_{13} q_1^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{14} q_1^{1/4} q_3^{3/4} + \frac{1}{4} \alpha_{15} q_1^{1/4} q_4^{3/4} + \frac{1}{4} \alpha_{16} q_1^{1/4} q_4^{3/4} + \frac{1}{4} \alpha_{23} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{24} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{25} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{26} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{27} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{28} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{29} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{30} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{31} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{32} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{33} q_1^{1/4} q_2^{1/4} q_3^{1/4} + \frac{1}{4} \alpha_{34} q_1^{1/4} q_2^{1/4} q_4^{1/4} + \frac{1}{4} \alpha_{35} q_1^{1/4} q_2^{1/4} q_3^{1/4} \right); \quad (39)$$
\[
\frac{x_2}{y} = \frac{1}{e^{q_2 t}} \left( \alpha_2 + \frac{1}{2} \alpha_5 q_1^{1/2} q_2^{-1/2} + \frac{1}{2} \alpha_8 q_2^{-1/2} q_3^{1/2} + \frac{1}{2} \alpha_9 q_2^{-1/2} q_4^{1/2} \\
+ \frac{1}{4} \alpha_{11} q_1^{3/4} q_2^{-3/4} + \frac{3}{4} \alpha_{12} q_1^{1/4} q_2^{-1/4} + \frac{3}{4} \alpha_{17} q_2^{-1/4} q_3^{3/4} \\
+ \frac{1}{4} \alpha_{18} q_3^{-3/4} q_3^{3/4} + \frac{3}{4} \alpha_{19} q_4^{-1/4} q_4^{1/4} + \frac{1}{4} \alpha_{20} q_4^{-3/4} q_4^{3/4} \\
+ \frac{1}{4} \alpha_{23} q_1^{1/2} q_2^{-3/4} q_3^{1/4} + \frac{1}{2} \alpha_{24} q_1^{1/4} q_2^{-1/2} q_3^{1/4} + \frac{1}{4} \alpha_{25} q_1^{1/4} q_2^{-3/4} q_3^{3/4} \\
+ \frac{1}{4} \alpha_{26} q_1^{1/2} q_2^{-3/4} q_4^{1/4} + \frac{1}{2} \alpha_{27} q_1^{1/4} q_2^{-1/2} q_4^{1/4} + \frac{1}{4} \alpha_{28} q_1^{1/4} q_2^{-3/4} q_4^{3/4} \\
+ \frac{1}{4} \alpha_{32} q_2^{-1/2} q_3^{3/4} q_4^{1/4} + \frac{1}{4} \alpha_{33} q_2^{-3/4} q_3^{1/2} q_4^{1/4} + \frac{1}{4} \alpha_{34} q_2^{-3/4} q_3^{3/4} q_4^{1/4} \\
+ \frac{1}{4} \alpha_{35} q_1^{1/4} q_2^{-3/4} q_3^{1/4} q_4^{1/4} \right); \tag{40}
\]

\[
\frac{x_3}{y} = \frac{1}{e^{q_2 t}} \left( \alpha_3 + \frac{1}{2} \alpha_6 q_1^{1/2} q_3^{-1/2} + \frac{1}{2} \alpha_8 q_2^{1/2} q_3^{-1/2} + \frac{1}{2} \alpha_9 q_3^{-1/2} q_4^{1/2} \\
+ \frac{1}{4} \alpha_{13} q_1^{3/4} q_3^{-3/4} + \frac{3}{4} \alpha_{14} q_1^{1/4} q_3^{-1/4} + \frac{1}{4} \alpha_{17} q_3^{-3/4} q_3^{3/4} \\
+ \frac{3}{4} \alpha_{18} q_2^{1/4} q_3^{-1/4} + \frac{3}{4} \alpha_{21} q_3^{-1/4} q_4^{1/4} + \frac{1}{4} \alpha_{22} q_3^{-3/4} q_4^{3/4} \\
+ \frac{1}{4} \alpha_{23} q_1^{1/2} q_2^{-3/4} q_3^{1/4} + \frac{1}{4} \alpha_{24} q_1^{1/4} q_2^{-1/2} q_3^{1/4} + \frac{1}{4} \alpha_{25} q_1^{1/4} q_2^{-3/4} q_3^{3/4} \\
+ \frac{1}{4} \alpha_{26} q_1^{1/2} q_2^{-3/4} q_4^{1/4} + \frac{1}{4} \alpha_{27} q_1^{1/4} q_2^{-1/2} q_4^{1/4} + \frac{1}{4} \alpha_{28} q_1^{1/4} q_2^{-3/4} q_4^{3/4} \\
+ \frac{1}{4} \alpha_{32} q_2^{-1/2} q_3^{3/4} q_4^{1/4} + \frac{1}{4} \alpha_{33} q_2^{-3/4} q_3^{1/2} q_4^{1/4} + \frac{1}{4} \alpha_{34} q_2^{-3/4} q_3^{3/4} q_4^{1/4} \\
+ \frac{1}{4} \alpha_{35} q_1^{1/4} q_2^{-3/4} q_3^{1/4} q_4^{1/4} \right); \tag{41}
\]
Concavity (in prices) requires that the Hessian matrix of the second derivatives of the cost function with respect to prices, $\nabla_{p p} C(p, y, t)$, is negative semidefinite. In practice, concavity of the cost function may not be satisfied. In this case, we impose concavity fully (at every data point in the sample) on the AIM model using methods suggested by Gallant and Golub (1984) in the case of the Fourier cost function and recently used in the context of consumer demand systems estimation by Serletis and Shahmoradi (2005) — we shall discuss this in detail in Section 4.2.

4 Data and Econometric Issues

We use annual KLEM (capital, labor, energy, and intermediate materials) data for total manufacturing in the United States over the period from 1953 to 2001. All series are from the website of the U.S. Bureau of Labor Statistics (BLS), at www.bls.gov/data/home.htm. 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 is to 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.

A major feature of the BLS data set is that constant returns to scale is built in by constructing input factor payments in such a way that they add up to the value of output. Thus, tests of returns to scale and scale bias are inappropriate, as are some tests of imperfect competition. Another feature of the BLS data set is that it provides the price and quantity series for purchased business services inputs. Directly collected data on purchased business
services are relatively scant, and for that reason they have been ignored by similar studies in the past. However, there is ample evidence of an increased use of purchased business services by industries over the post-war period and there are two important issues to consider. The first is that a sizable and growing input should not be ignored in productivity measurement, if aggregate inputs are not to be underestimated and mismeasured. The other is the possibility of substitution between capital, labor, and services purchased from outside. Examples of the latter are the substitution of leased equipment for owned capital and purchased accounting for services performed by payroll employees.

Maximum likelihood estimates of the three locally flexible cost functions with or without curvature imposed is straightforward, and can be approached in a variety of well-known ways. The estimation of the AIM model without curvature imposed can also be approached easily in the same way as with locally flexible cost functions.

The estimation, however, of the AIM model with curvature imposed cannot be approached in the usual way, and has to resort to some more advanced methods. For example, Gallant and Golub (1984) used the NPSOL subroutine of the Stanford Systems Optimization Laboratory to estimate the constrained Fourier cost function without technical change. Also Barnett et al. (1991) used numerical Bayesian estimation to solve a relatively simple constrained AIM cost function with only two factors (capital and labor) and no technical change. In this paper, we follow Gallant and Golub (1984) and Serletis and Shahmoradi (2005) and use the TOMLAB/NPSOL tool box with MATLAB — see http://tomlab.biz/products/npsol. 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.

4.1 Parametric Estimation of the Locally Flexible Forms

In order to estimate equation systems such as (13), (19), and (27), a stochastic component, $\epsilon_t$, is added to the set of input-output equations or share equations as follows

$$w_t = \psi(p_t, y, t, \theta) + \epsilon_t,$$

where $w = (w_1, \ldots, w_n)'$ is the vector of input-output ratios in the case of the GL and NQ models and that of input shares in the case of the translog model. $\epsilon_t$ is a vector of stochastic errors and we assume that $\epsilon \sim N(0, \Omega)$ where $0$ is a null matrix and $\Omega$ is the $n \times n$ symmetric positive definite error covariance matrix. $\psi(p_t, y, t, \theta) = (\psi_1(p_t, y, t, \theta), \ldots, \psi_n(p_t, y, t, \theta))'$, and $\psi_i(p_t, y, t, \theta)$ is given by the right-hand side of each of (13), (19), and (27).

In the case of the translog model, since the shares in (19) sum to unity, the random disturbances corresponding to the four share equations sum to zero and 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 and the parameter estimates of the deleted equation can be recovered from the restrictions imposed.

Another issue concerning our stochastic specification is that of endogeneity. At the individual firm level, it may be reasonably assumed that inputs prices on the right hand side of (43) are exogenous. At the more aggregated industry level (like U.S. manufacturing), however, input prices are less likely to be exogenous. In this literature, the possibility of endogeneity has been addressed by using iterative three-stage least squares (3SLS), but the results generally have been about the same as those with iterative Zellner estimation — see, for example, Barnett et al. (1991). Diewert and Fox (2004) also argue that instrumental variables estimation may be more biased, since the instruments may not be completely exogenous, and Burnside (1996) shows that results can vary markedly depending on the set of instruments used. In this paper, we choose to use the more commonly used iterative Zellner method of estimation.

The estimation is performed in TSP/GiveWin (version 4.5), using the LSQ procedure, and the regularity conditions are checked as follows:

- Positivity is checked by checking if the estimated cost is positive,
  \[ C(p, y, t) > 0. \]

- Monotonicity is checked by direct computation of the values of the first gradient vector of the estimated cost function with respect to \( p \). It is satisfied if \( \nabla_p C(p, y, t) > 0 \).

- Curvature requires the Hessian matrix of the cost function to be negative semidefinite and is checked by performing a Cholesky factorization of that matrix and checking whether the Cholesky values are nonpositive [since a matrix is negative semidefinite if its Cholesky factors are nonpositive — see Lau (1978, Theorem 3.2)]. Curvature can also be checked by examining the eigenvalues of the Hessian matrix provided that the monotonicity condition holds. It requires that these eigenvalues be negative or zero.

### 4.2 Semi-Nonparametric Estimation of the AIM(2) Cost Function

The AIM(2) factor demand system can be written as

\[ z_t = \psi(p, y, t, \theta) + \epsilon_t, \tag{44} \]

where \( z = (z_1, \ldots, z_n)' \) is the vector of input-output ratios, \( \theta = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, \vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4) \), and \( \psi(p, y, t, \theta) \) is given by the the right hand side of (39)-(42).

As Gallant and Golub (1984, p. 298) put it,
“all statistical estimation procedures that are commonly used in econometric research can be formulated as an optimization problem of the following type [Burguete, Gallant and Souza (1982)]

\[ \hat{\theta} \text{ minimizes } \varphi(\theta) \text{ over } \Theta \]  

(45)

with \( \varphi(\theta) \) twice continuously differentiable in \( \theta \).”

Notice that \( \psi(p, y, t, \theta) \) is nonlinear in \( \theta_1, \theta_2, \theta_3, \) and \( \theta_4 \), and therefore the AIM(2) factor demand system in (44) can be fitted using Gallant’s (1975, p. 36) seemingly unrelated nonlinear regression method to estimate \( \theta \). Hence, \( \varphi(\theta) \) has the form

\[ \varphi(\theta) = \frac{1}{T} \epsilon_t' \epsilon_t = \frac{1}{T} \sum_{t=1}^{T} (z_t - \psi(\cdot))' \hat{\Omega}^{-1} (z_t - \psi(\cdot)), \]  

(46)

where \( \hat{\Omega} \) is an estimate of the error variance-covariance matrix of (44). In minimizing (44), we use the TOMLAB/NPSOL tool box with MATLAB. 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.

We first run an unconstrained optimization using (45). As results in nonlinear optimization are sensitive to the initial parameter values, to achieve global convergence, we randomly generated 500 sets of initial parameter values and chose the starting \( \theta \) that led to the lowest value of the objective function. We also check the regularity conditions, i.e. positivity, monotonicity, and curvature conditions, using the same methods as specified above for the three locally flexible functional forms.

In case where the curvature conditions are not satisfied at all observations, we then use the NPSOL nonlinear programming program to minimize \( \varphi(\theta) \) subject to the constraint that the four eigenvalues of the Hessian matrix, \( H \), are non-positive. This is because a necessary and sufficient condition for the concavity of \( H \) is that all its eigenvalues are nonpositive — see, for example, Morey (1986). The first derivatives of these eigenvalues are needed for the optimization algorithms and can be easily obtained using Matlab’s Symbolic Math Toolbox. Thus, our constrained optimization problem can also be written as

\[ \min_{\theta} \varphi(\theta) \text{ subject to } \varphi_i(p, y, t, \theta) < 0, \quad i = 1, \ldots, n, \]

where \( \varphi_i(p, y, t, \theta), i = 1, \ldots, n, \) are the eigenvalues of the Hessian matrix of the AIM(2) cost function. With the constrained optimization method, we can impose curvature restrictions at any arbitrary set of points — at a single data point, over a region of data points, or fully (at every data point in the sample).
5 Empirical Evidence

5.1 Economic Regularity

Tables 1-4 contain a summary of results from the GL, translog, NQ, and AIM(2) models in terms of parameter estimates and theoretical regularity violations when the models are estimated without the curvature conditions imposed and with the curvature conditions imposed. Clearly, all models satisfy positivity and monotonicity at all sample observations when curvature is not imposed. However, all three locally flexible models — the GL, translog, and NQ — violate curvature at all 49 observations when curvature conditions are not imposed. Similarly, the AIM(2) model violates curvature at 33 data points when curvature is not imposed.

Because regularity hasn’t been attained for any of the models, we follow the procedures discussed in Section 3 to impose curvature. In the case of the GL and translog models, we impose local curvature using the Ryan and Wales (2000) procedure. However, as noted by Ryan and Wales (2000), the ability of locally flexible models to satisfy curvature at sample observations other than the point of approximation, depends on the choice of the approximation point. Thus, we estimate each model 49 times (a number of times equal to the number of observations) and report results for the best approximation point (best in the sense of satisfying the curvature conditions at the largest number of observations) — the best approximation point is 1982 for the GL and 1981 for the translog. In the case of the NQ model we impose global curvature following the procedure suggested by Diewert and Wales (1987). As for the AIM(2) model, we minimize $\varphi(\theta)$ subject to the constraint that the cost function is locally concave in 1981 and also subject to the constraint that it is fully concave (concave at every data point).

The estimation results of the three locally flexible functional forms with curvature imposed are reported in the second column of Tables 1-3. Our findings in terms of regularity violations when the curvature conditions are imposed are disappointing in the case of the GL and translog models. In particular, the imposition of local curvature on the translog model reduces the number of curvature violations from 49 to 6. The performance of the GL is not satisfactory either, since the imposition of local curvature does not completely eliminate the curvature violations; it reduces the number of curvature violations from 49 to 2. As Barnett (2002, p.199) put it, without satisfaction of all three theoretical regularity conditions, “the second-order conditions for optimizing behavior fail, and duality theory fails. The resulting first-order conditions, demand functions, and supply functions become invalid.” As expected, however, the imposition of global curvature (at all possible prices) on the NQ model reduces the number of curvature violations to zero, without any induced violations of monotonicity.

Using NPSOL we imposed the curvature condition on the AIM(2) model and report the results in the second and third columns of Table 4 — the second column shows the results when the curvature constraint is imposed locally (in 1981) and the third column shows the
results when the constraint is imposed at every data point in the sample. Clearly, the effect of imposing the curvature constraint locally is negligible, as the number of curvature violations drops only from 33 to 32. However, the imposition of the curvature constraint at every data point in the sample has a significant impact on the AIM(2) model, as we obtain parameter estimates that are consistent with all three theoretical regularity conditions, at every data point in the sample; that is, fully.

5.2 Econometric Regularity

We have estimated input-output demand equations and share equations from aggregate time series data and highlighted the challenge inherent with achieving economic regularity and the need for economic theory to inform econometric research. Incorporating restrictions from economic theory seems to be gaining popularity as there are also numerous recent papers that estimate stochastic dynamic general equilibrium models using economic restrictions — see, for example, Aliprantis et al. (2007). With the focus on economic theory, however, we have ignored econometric regularity. In particular, we have ignored unit root and cointegration issues, because the combination of nonstationary data and nonlinear estimation in large models like the ones in this paper is an extremely difficult problem.

In this regard, it should be noted that we used two alternative unit root testing procedures — the augmented Dickey-Fuller (ADF) test [see Dickey and Fuller (1981)] and the non-parametric, \(Z(t_\alpha)\) test of Phillips (1987) and Phillips and Perron (1987) — to deal with anomalies that arise when the data are not very informative about whether or not there is a unit root, and found that our input-output ratios, budget shares, and price variables are all integrated of order one [or I(1) in the terminology of Engle and Granger (1987)]. It follows then that for input-output demand equations and share equations to make any sense the variables must be cointegrated in levels; that is, the equation errors must be stationary. However, unit root test results on the residuals of the locally flexible systems — the generalized Leontief, translog, and normalized quadratic models — and the globally flexible AIM model indicate that they are nonstationary.

If the errors are nonstationary, then there is no theory linking the left hand side to the right hand side variables in equation (43) or, equivalently, no evidence for the theoretical models in level form. In such cases, some important nonstationary variables might have been omitted. Allowing for first order serial correlation, as is usually done in the literature, is almost the same as taking first differences of the data if the autocorrelation coefficient is close to unity. In that case, the equation errors become stationary, but there is no theory for the models in first differences. Moreover, as argued by Serletis and Shahmoradi (2007), serial correlation correction increases the number of curvature violations and also leads to induced violations of monotonicity.

It is also to be noted that even if the errors are stationary and the estimates are super consistent, as argued by Attfield (1997) and Ng (1995), standard estimation procedures are
inadequate for obtaining correctly estimated standard errors for coefficients in cointegrating equations. In that case, if the equations were all linear, the DOLS method of Stock and Watson (1993) or the FM-OLS method of Phillips (1995) could have been used to obtain correctly estimated standard errors. With our nonlinear models, however, some sort of modification of these procedures is called for, but this is a very difficult issue to deal with.

With the generalized Leontief and translog models failing both economic and econometric regularity and the NQ and AIM(2) models failing econometric regularity, in what follows we report total factor productivity estimates and elasticity estimates based only on the NQ and AIM(2) models.

5.3 Total Factor Productivity Trends

Figure 1 provides year-by-year total factor productivity estimates with the NQ and AIM(2) models, together with productivity measures formed from the Fisher ideal index and the smoothed Fisher ideal index. Roughly speaking, the total factor productivity estimates from the NQ and AIM(2) models exhibit similar patterns. First, both of them show a general tendency to rise over the sample period. In particular, the estimates based on the NQ model rise markedly from 0.55% to 1.99% over the sample period whereas those based on the AIM(2) model rise moderately from 0.91% to 1.16%. Second, the two models have produced average total factor productivity measures which are very close to each other. In particular, the average total factor productivity from the NQ model is 1.08%, compared with 1.02% from the AIM(2) model.

To further evaluate the performance of the NQ and the AIM(2) models in capturing technical change, we calculate the productivity growth in U.S. manufacturing as a benchmark, using the Fisher ideal index. We first calculate the Fisher ideal quantity index for the four inputs as

$$g^t = \left[ \frac{\sum_{j=1}^{n} p_j^t x_j^t \sum_{j=1}^{n} p_j^{t-1} x_j^{t-1}}{\sum_{j=1}^{n} p_j^t x_j^t \sum_{j=1}^{n} p_j^{t-1} x_j^{t-1}} \right]^{1/2}$$

and then calculate the quantity index for the single output as $G^t = y^t / y^{t-1}$. The Fisher ideal total factor productivity index is then obtained as

$$\frac{G^t}{g^t} - 1$$

Following Fox (1996), we also obtain a smoothed Fisher ideal total factor productivity index by regressing the raw Fisher ideal series on a constant and a time trend and calculating the fitted values. Both of these indexes are plotted in Figure 1. Clearly, both the NQ and AIM(2) measures pass close by the mean of the raw Fisher ideal index series, which is volatile from year to year. Further, both of them evolve in a similar pattern as the smoothed Fisher
ideal index which also shows a general tendency to rise from 0.98% to 1.31% over the sample period. In this sense, the productivity growth measures from both the NQ and AIM(2) models can be regarded as smoothed versions of that from the Fisher ideal index. Generally speaking, both the NQ and AIM(2) models perform pretty well in modelling productivity growth in the U.S. manufacturing industry. However, a close look at Figure 1 reveals that the AIM(2) measure resembles the curve of the smoothed Fisher ideal index more closely. Moreover, the AIM(2) model captures (though not noticeably) the slowdown in productivity between 1974 and 1994, which is missed by the NQ model.

An advantage of the AIM(2) model over the NQ model is that total factor productivity can be easily decomposed into growth rates of input-specific efficiencies ($\vartheta'$s). In particular, substituting the AIM(2) cost function into (6), we can obtain the specific total factor productivity formula for the AIM(2) model

$$\text{TFP}_{\text{AIM}} = \sum_{i=1}^{n} s_i \vartheta_i.$$  \hspace{1cm} (47)

Equation (47) shows that total factor productivity estimates based on the AIM(2) model are an input cost-share weighted average of the growth rates of factor efficiencies. As shown in Table 4, the long-run growth rate of the efficiency of capital, $\vartheta_1$, is 2.94% per year. For labor, energy, and materials, the long-run growth rates of efficiency are found to be 0.04% ($\vartheta_2$), 5.47% ($\vartheta_3$), and 0.96 % ($\vartheta_4$), respectively. We further define

$$CT_i = \frac{s_i \vartheta_i}{\text{TFP}_{\text{AIM}}}$$

to be the contribution of factor $i$ to total factor productivity, and plot in Figure 2 the contribution of each factor to total factor productivity. Clearly, capital and materials have been the dominant factors causing productivity growth, with energy having a moderate positive contribution to total factor productivity due to its small input cost share. Labor has a positive and small impact on total factor productivity.

### 5.4 Elasticity Estimates

We begin by presenting the own- and cross-price elasticities in Table 5, evaluated at the mean of the data. The signs of all own-price elasticities, $\eta_{ii}$, appear reasonable for both models since they are all negative (as predicted by the theory), with the absolute values being less than 1, indicating that the demands for all four inputs are inelastic. However, the AIM(2) model shows larger own-price elasticities in absolute value than the NQ model. In particular, $\eta_{KK}$ from the AIM(2) is -0.522 which is about twice as large as that from the NQ model. Similarly, $\eta_{LL}$ (-0.592), $\eta_{EE}$ (-1.927), and $\eta_{MM}$ (-0.297) from the AIM(2) model are about 3-10 times as large as their counterparts from the NQ model. This implies that
capital, labor, energy, and materials are all more responsive to their own prices according to the AIM(2) than according to the NQ model.

We believe there are actually good reasons to graph the own-price elasticities that we have estimated. Figures 3-6 present the own-price elasticities for $K$, $L$, $E$, and $M$ for each observation. Clearly, the own-price elasticities are negative at all data points for both models, as predicted by the theory. A prominent difference, as can be seen from these figures, is that the own-price elasticities from the AIM(2) model show quite large variations, whereas those from the NQ trend over time. This problem of lacking variations in own-price elasticities over time with the NQ model was first noted by Dievert and Lawrence (2002) and referred to by them as ‘the problem of trending elasticities.’ As can seen below, this problem in the NQ model is also reflected in its cross elasticities and carried over to its Morishima elasticities. To cure this problem with the NQ model, Dievert and Lawrence (2002) suggested imposing flexibility at two sample points.

As with the own-price elasticities, the cross-price elasticities differ significantly between the two models (see Table 5). Moreover, results not presented here, but available upon request, indicate that the cross-price elasticities from the AIM(2) model vary considerably over time whereas those from the NQ are very stable. For example, while the AIM(2) model shows the capital-labor substitution ($\eta_{KL}$) in the range between 0.43 to 0.93, the NQ model show a very stable $\eta_{KL}$, varying in a rather small range between 0.20 to 0.24. In addition, the two models show different relations between some of the four inputs. For example, both models classify capital and labor (see $\eta_{KL}$ and $\eta_{LK}$) and energy and materials (see $\eta_{EM}$ and $\eta_{ME}$) as substitutes, but are inconsistent in their classification of capital and materials (see $\eta_{KM}$ and $\eta_{MK}$) and labor and materials (see $\eta_{LM}$ and $\eta_{ML}$).

We now turn to the estimates of the Morishima elasticities of substitution, $\sigma_{ij}^m$ ($i,j = K, L, E, and M$), presented in Figures 7-18. Since the Morishima elasticities of substitution are just a simple function of related own- and cross-price elasticities (see equation (10)), the differences in own- and cross-price elasticities between the NQ and AIM(2) models also show up in the Morishima elasticities of substitution. In particular, the Morishima elasticities of substitution from the AIM(2) model vary considerably whereas those from the NQ model are very stable over the sample period. Moreover, the Morishima elasticities of substitution from the AIM(2) model are generally larger than the corresponding ones from the NQ model. Again, we are more interested in the Morishima elasticities of substitution obtained from the AIM(2) model and discuss them in more detail in what follows.

Let’s consider first the Morishima elasticity of substitution between $K$ and $L$, $\sigma_{KL}^m$, which represents the percentage change in the capital services to the labor quantity ratio, $K/L$, when the relative price $P_L/P_K$ is changed by changing $P_L$ and holding $P_K$ constant. Figures 7 and 10 reveal that at each data point, $\sigma_{KL}^m > 0$, and the average estimated $\sigma_{KL}^m$ is 1.303, compared with an average estimated $\sigma_{LK}^m$ of 0.766. Thus, capital services and labor are Morishima substitutes, irrespective of whether the price of labor or the price of capital services changes.
Of particular interest are $\sigma_{KE}^m$ and $\sigma_{EK}^m$ in Figures 8 and 13. The estimates of $\sigma_{KE}^m$ are positive, but $\sigma_{EK}^m$ is positive for most of the sample period (in particular, from 1960 to 1998) and negative from 1999 to 2001. Thus, capital services and energy are always Morishima substitutes when the price of energy changes, but can be either Morishima complements (as from 1960 to 1998) or Morishima substitutes (as for the rest of the sample period) when the price of capital services changes. In other words, an increase in the price of energy (holding the price of capital services constant) always leads to an increase in the $K/E$ ratio, but an increase in the price of capital services (holding the price of energy constant) can lead to either an increase or a decrease in the $E/K$ ratio. We also notice that at each data point between 1960 and 1998, when both $\sigma_{KE}^m$ and $\sigma_{EK}^m$ are positive, $\sigma_{KE}^m$ is greater than $\sigma_{EK}^m$, and the average estimated $\sigma_{KE}^m$ is 1.926 whereas the average estimated $\sigma_{EK}^m$ is 0.247.

Next we consider $\sigma_{LE}^m$ and $\sigma_{EL}^m$ — see Figures 11 and 14. $\sigma_{LE}^m$ is positive throughout, but $\sigma_{EL}^m$ is negative prior to 1977 and positive afterwards. Thus, labor and energy are always Morishima substitutes when energy prices change, but they can be either Morishima complements or substitutes when the price of labor changes. The estimated $\sigma_{KM}^m$ and $\sigma_{MK}^m$ are always positive — see Figures 9 and 16. Thus, capital services and materials are Morishima substitutes irrespective of whether the price of materials changes or the price of capital services changes. The average estimated $\sigma_{KM}^m$ is 0.221, compared with an average estimated $\sigma_{MK}^m$ of 0.390. Similarly, the estimates of $\sigma_{LM}^m$ and $\sigma_{ML}^m$ are positive throughout (see Figures 12 and 17), indicating that labor and materials are Morishima substitutes irrespective of whether the price of materials changes or the price of labor changes. The average estimated $\sigma_{LM}^m$ is 0.720, compared with an average estimated $\sigma_{ML}^m$ of 0.967. Finally, the estimates of $\sigma_{EM}^m$ are positive and those of $\sigma_{ME}^m$ are positive for most of the sample period, but negative after 1994.

6 Conclusion

We have investigated productivity issues in the U.S. (total) manufacturing industry, in the context of three popular locally flexible functional forms — the generalized Leontief (GL), translog, and normalized quadratic (NQ) — and one globally flexible functional form — the Asymptotically Ideal Production Model (AIM). In doing so, we have extended the Barnett et al. (1991) AIM model, by incorporating (for the first time in the literature) technical change through the factor-augmenting efficiency index approach, proposed by Thomsen (2000).

We estimated the three locally flexible functional forms parametrically and the globally flexible functional form semi-nonparametrically and treated the curvature property as a maintained hypothesis. In particular, we imposed local curvature on the GL and translog models using procedures suggested by Ryan and Wales (2000), we imposed global curvature on the NQ using procedures suggested by Diewert and Wales (1987), and imposed local and global curvature on the AIM(2) model using procedures suggested by Gallant and Golub.
(1984) and more recently by Serletis and Shahmoradi (2005). We also showed that (with our data set) the imposition of local curvature does not always assure theoretical regularity, because of curvature violations at other points within the region of the data. We believe that this is a typical result in the literature that uses locally flexible functional forms and alert researchers to the kinds of problems that arise when all three theoretical regularity conditions are not satisfied — see also Barnett (2002) and Barnett and Pasupathy (2003).

We provided a comparison between the NQ and AIM cost functions, the only two models that satisfy all three theoretical regularity conditions. We found that the AIM(2) cost function with technical change introduced through the factor-augmenting efficiency index approach performs better than traditional locally flexible function forms and gives more accurate estimates of total factor productivity. We also found that the elasticities from the AIM(2) model are generally larger and show more variation than those from the NQ model, which is consistent with Gallant and Golub (1984) who employed a different globally flexible functional form — the Fourier. Finally, we discussed the elasticities based on the AIM(2) model to shed some new light on the substitutability/complementarity relationship between capital, labor, energy, and materials.

Although we have achieved economic regularity (in terms of curvature, positivity, and monotonicity) with the NQ and AIM(2) models, we have not achieved econometric regularity (in terms of stationary equation errors), which makes interpreting our results difficult. Moreover, our econometric modeling assumes a serially uncorrelated Gaussian measurement error, or equivalently serially independent measurement error. To simultaneously achieve both economic and econometric regularity seems to be a challenging task and an area for potentially productive future research. It could also be that the econometric irregularity is caused by our treatment of technical change. That is, we treated technical change as being smooth over the sample period, but there are fairly large year to year fluctuations in technical change as well as secular trends in total factor productivity growth. Using the spline techniques pioneered by Diewert and Wales (1993) and Fox (1996) to model these trends in the context of the production models used in this paper is work that we are currently undertaking.

Finally, the Bayesian approach, pioneered by Terrell (1996) and Griffiths et al. (2000) in imposing regularity on linear factor demand systems, could also be directly used to estimate the first three locally flexible cost models presented in this paper, since all these three models are linear. For our AIM model with technical change, the application of Bayesian inference is more complicated due to its highly nonlinear nature and also the difficulty in finding reasonable informative priors. The Griffiths and Chotikapanich (1997) method can be used in this case after some appropriate modifications. The Bayesian approach has two major advantages that traditional econometric methods commonly used for productivity estimation do not possess. First, the Bayesian approach provides exact (small-sample) inference on the productivity components (i.e. technical change, efficiency change, and returns to scale) which in many cases are nonlinear functions of estimated parameters, whereas the
traditional methods provide only point estimates of these productivity components without statistical inference. Second, and even more importantly, the Bayesian approach allows us to incorporate the theoretical regularity restrictions of neoclassical microeconomic theory in the estimation. This can be done either by using the accept-reject algorithm — see Terrel (1996) — or the Metropolis-Hastings algorithm — see Griffiths et al. (2000).
References


Table 1

**Generalized Leontief Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unrestricted</th>
<th>Local curvature imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{11}$</td>
<td>.0966 (.000)</td>
<td>.0884 (.000)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>.0438 (.017)</td>
<td>.0844 (.000)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>$-0.082$ (.024)</td>
<td>$-0.0101$ (.000)</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>.0633 (.000)</td>
<td>.0507 (.000)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>.5108 (.000)</td>
<td>.3130 (.000)</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>.0006 (.925)</td>
<td>.0163 (.023)</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>$-0.1371$ (.000)</td>
<td>$-0.0386$ (.000)</td>
</tr>
<tr>
<td>$\beta_{33}$</td>
<td>.0650 (.054)</td>
<td>.0104 (.007)</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>.0264 (.001)</td>
<td>.0123 (.059)</td>
</tr>
<tr>
<td>$\beta_{44}$</td>
<td>.4033 (.000)</td>
<td>.3044 (.000)</td>
</tr>
<tr>
<td>$\beta_{1t}$</td>
<td>.0003 (.274)</td>
<td>.0001 (.045)</td>
</tr>
<tr>
<td>$\beta_{2t}$</td>
<td>$-0.0073$ (.000)</td>
<td>$-0.0052$ (.000)</td>
</tr>
<tr>
<td>$\beta_{3t}$</td>
<td>$-0.0001$ (.430)</td>
<td>.0001 (.005)</td>
</tr>
<tr>
<td>$\beta_{4t}$</td>
<td>.0026 (.000)</td>
<td>.0011 (.000)</td>
</tr>
</tbody>
</table>

Positivity violations 0 0
Monotonicity violations 0 0
Curvature violations 49 2

Notes: Sample period, annual data 1953-2001 ($T = 49$).
Table 2

TRANSLOG PARAMETER ESTIMATES

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unrestricted</th>
<th>Local curvature imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-.0512 (.000)</td>
<td>1.3182 (.000)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>.1635 (.000)</td>
<td>.2546 (.000)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>.4580 (.000)</td>
<td>.2568 (.000)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>.0268 (.000)</td>
<td>.2265 (.000)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>.1055 (.000)</td>
<td>.1176 (.000)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-.0553 (.000)</td>
<td>-.0747 (.000)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-.0097 (.000)</td>
<td>-.0113 (.216)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>.2467 (.000)</td>
<td>.1810 (.000)</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>-.0175 (.000)</td>
<td>-.0575 (.000)</td>
</tr>
<tr>
<td>$\beta_{33}$</td>
<td>.0171 (.000)</td>
<td>.0886 (.000)</td>
</tr>
<tr>
<td>$\beta_{1t}$</td>
<td>.0024 (.000)</td>
<td>.0031 (.000)</td>
</tr>
<tr>
<td>$\beta_{2t}$</td>
<td>-.0071 (.000)</td>
<td>-.0050 (.000)</td>
</tr>
<tr>
<td>$\beta_{3t}$</td>
<td>.0003 (.015)</td>
<td>.0009 (.001)</td>
</tr>
<tr>
<td>$\beta_t$</td>
<td>-.0026 (.006)</td>
<td>-.0075 (.000)</td>
</tr>
<tr>
<td>$\beta_{tt}$</td>
<td>-.0001 (.002)</td>
<td>-.00002 (.550)</td>
</tr>
</tbody>
</table>

Positivity violations 0 0
Monotonicity violations 0 0
Curvature violations 49 6

Notes: Sample period, annual data 1953-2001 ($T = 49$).
Table 3

NQ Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unrestricted</th>
<th>Global curvature imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>.1657 (.000)</td>
<td>.1677 (.000)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>.4179 (.000)</td>
<td>.4251 (.000)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>.0257 (.000)</td>
<td>.0254 (.000)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>.3542 (.000)</td>
<td>.3556 (.000)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>.0287 (.000)</td>
<td>.0331 (.003)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-.0057 (.005)</td>
<td>-.0059 (.015)</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>.0409 (.000)</td>
<td>.0324 (.000)</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>.0006 (.861)</td>
<td>.0059 (.126)</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>-.0868 (.000)</td>
<td>-.0197 (.199)</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>.0136 (.000)</td>
<td>.0100 (.002)</td>
</tr>
<tr>
<td>$\beta_{1t}$</td>
<td>.0006 (.074)</td>
<td>.0001 (.686)</td>
</tr>
<tr>
<td>$\beta_{2t}$</td>
<td>-.0078 (.000)</td>
<td>-.0062 (.000)</td>
</tr>
<tr>
<td>$\beta_{3t}$</td>
<td>-.0001 (.391)</td>
<td>-.0002 (.039)</td>
</tr>
<tr>
<td>$\beta_{4t}$</td>
<td>.0033 (.000)</td>
<td>.0010 (.082)</td>
</tr>
</tbody>
</table>

Positivity violations: 0
Monotonicity violations: 0
Curvature violations: 49

Notes: Sample period, annual data 1953-2001 ($T = 49$).
### Table 4. AIM(2) Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unconstrained estimates</th>
<th>Curvature constrained at 1981 fully</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_1$</td>
<td>$0.0072$</td>
<td>$0.0071$</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>$0.0154$</td>
<td>$0.0154$</td>
</tr>
<tr>
<td>$\psi_3$</td>
<td>$0.0079$</td>
<td>$0.0079$</td>
</tr>
<tr>
<td>$\psi_4$</td>
<td>$0.0048$</td>
<td>$0.0049$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$36.5369$</td>
<td>$36.5660$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$48.0633$</td>
<td>$48.4436$</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-7.5918$</td>
<td>$-7.6265$</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>$57.7529$</td>
<td>$57.6138$</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>$132.6923$</td>
<td>$133.0500$</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>$12.8664$</td>
<td>$12.7332$</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>$226.2107$</td>
<td>$225.7530$</td>
</tr>
<tr>
<td>$\alpha_8$</td>
<td>$224.7218$</td>
<td>$224.7091$</td>
</tr>
<tr>
<td>$\alpha_9$</td>
<td>$-98.4541$</td>
<td>$-98.3085$</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>$-99.6927$</td>
<td>$-99.4859$</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>$-73.6204$</td>
<td>$-73.8018$</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>$-148.9083$</td>
<td>$-149.3250$</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>$46.2347$</td>
<td>$46.1778$</td>
</tr>
<tr>
<td>$\alpha_{14}$</td>
<td>$-3.6166$</td>
<td>$-3.6960$</td>
</tr>
<tr>
<td>$\alpha_{15}$</td>
<td>$-108.7704$</td>
<td>$-108.6506$</td>
</tr>
<tr>
<td>$\alpha_{16}$$\alpha_{16}$</td>
<td>$-134.5830$</td>
<td>$-134.1460$</td>
</tr>
<tr>
<td>$\alpha_{17}$</td>
<td>$-158.5638$</td>
<td>$-159.1355$</td>
</tr>
<tr>
<td>$\alpha_{18}$</td>
<td>$-51.6272$</td>
<td>$-51.4163$</td>
</tr>
<tr>
<td>$\alpha_{19}$</td>
<td>$96.5016$</td>
<td>$95.9014$</td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>$-45.5792$</td>
<td>$-45.6127$</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>$90.5297$</td>
<td>$90.5274$</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>$-51.3034$</td>
<td>$-51.1016$</td>
</tr>
<tr>
<td>$\alpha_{23}$</td>
<td>$-46.7539$</td>
<td>$-46.9368$</td>
</tr>
<tr>
<td>$\alpha_{24}$</td>
<td>$182.7781$</td>
<td>$183.1389$</td>
</tr>
<tr>
<td>$\alpha_{25}$</td>
<td>$-133.1851$</td>
<td>$-133.0411$</td>
</tr>
<tr>
<td>$\alpha_{26}$</td>
<td>$-21.2014$</td>
<td>$-21.2212$</td>
</tr>
<tr>
<td>$\alpha_{27}$</td>
<td>$38.3530$</td>
<td>$38.5663$</td>
</tr>
<tr>
<td>$\alpha_{28}$</td>
<td>$-15.7660$</td>
<td>$-15.6267$</td>
</tr>
<tr>
<td>$\alpha_{29}$</td>
<td>$-113.1907$</td>
<td>$-112.5742$</td>
</tr>
<tr>
<td>$\alpha_{30}$</td>
<td>$127.3465$</td>
<td>$127.6929$</td>
</tr>
<tr>
<td>$\alpha_{31}$</td>
<td>$-9.9053$</td>
<td>$-10.5092$</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>$-132.7490$</td>
<td>$-131.3271$</td>
</tr>
<tr>
<td>$\alpha_{33}$</td>
<td>$-191.5419$</td>
<td>$-192.2739$</td>
</tr>
<tr>
<td>$\alpha_{34}$</td>
<td>$347.8034$</td>
<td>$347.3930$</td>
</tr>
<tr>
<td>$\alpha_{35}$</td>
<td>$-20.7763$</td>
<td>$-21.4390$</td>
</tr>
</tbody>
</table>

$S(\hat{\theta})$ 0.0070  0.0071  0.0103

Positivity violations 0  0  0
Monotonicity violations 0  0  0
Curvature violations 33  32  0

Note: Sample period, annual data 1953-2001 ($T = 49$).
Figure 1. Total Factor Productivity Estimates
Figure 2. Factor Contributions to Total Factor Productivity

- Capital
- Labor
- Energy
- Materials

# Table 5

**Price Elasticities at the Mean**

<table>
<thead>
<tr>
<th>Factor $i$</th>
<th>Model</th>
<th>$\eta_{iK}$</th>
<th>$\eta_{iL}$</th>
<th>$\eta_{iE}$</th>
<th>$\eta_{iM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K)</td>
<td>NQ</td>
<td>-.267</td>
<td>.216</td>
<td>-.040</td>
<td>.092</td>
</tr>
<tr>
<td></td>
<td>AIM(2)</td>
<td>-.522</td>
<td>.804</td>
<td>.032</td>
<td>-.314</td>
</tr>
<tr>
<td>(L)</td>
<td>NQ</td>
<td>.080</td>
<td>-.071</td>
<td>.026</td>
<td>-.035</td>
</tr>
<tr>
<td></td>
<td>AIM(2)</td>
<td>.314</td>
<td>-.592</td>
<td>-.011</td>
<td>.289</td>
</tr>
<tr>
<td>(E)</td>
<td>NQ</td>
<td>-.239</td>
<td>.409</td>
<td>-.547</td>
<td>.376</td>
</tr>
<tr>
<td></td>
<td>AIM(2)</td>
<td>.264</td>
<td>-.249</td>
<td>-1.927</td>
<td>1.912</td>
</tr>
<tr>
<td>(M)</td>
<td>NQ</td>
<td>.042</td>
<td>-.044</td>
<td>.030</td>
<td>-.028</td>
</tr>
<tr>
<td></td>
<td>AIM(2)</td>
<td>-.142</td>
<td>.335</td>
<td>.104</td>
<td>-.297</td>
</tr>
</tbody>
</table>

Note: Sample period, annual data 1953-2001 ($T = 49$).
Figure 3. Own Price Elasticities for Capital
Figure 4. Own Price Elasticities for Labor
Figure 5. Own Price Elasticities for Energy
Figure 6. Own Price Elasticities for Materials

Year


AIM(2)

NQ
Figure 7. Morishima Elasticities of Substitution between K and L with the Price of L Changing
Figure 8. Morishima Elasticities of Substitution between $K$ and $E$ with the Price of $E$ Changing.
Figure 9. Morishima Elasticities of Substitution between $K$ and $M$ with the Price of $M$ Changing
Figure 10. Morishima Elasticities of Substitution between $L$ and $K$ with the Price of $K$ Changing
Figure 11. Morishima Elasticities of Substitution between $L$ and $E$ with the Price of $E$ Changing
Figure 12. Morishima Elasticities of Substitution between $L$ and $M$ with the Price of $M$ Changing
Figure 13. Morishima Elasticities of Substitution between $E$ and $K$ with the Price of $K$ Changing
Figure 14. Morishima Elasticities of Substitution between $E$ and $L$ with the Price of $E$ Changing
Figure 15. Morishima Elasticities of Substitution between $E$ and $M$ with the Price of $M$ Changing
Figure 16. Morishima Elasticities of Substitution between $M$ and $K$ with the Price of $K$ Changing
Figure 17. Morishima Elasticities of Substitution between $M$ and $L$ with the Price of $L$ Changing
Figure 18. Morishima Elasticities of Substitution between $M$ and $E$ with the Price of $E$ Changing