Total Factor Productivity Computed and Evaluated Using Multi-Step Perturbation[*]

Baoline Chen
Bureau of Economic Analysis
1441 L Street, NW
Washington, DC 20230
e-mail: baoline.chen@bea.gov

and

Peter A. Zadrozny
Bureau of Labor Statistics
2 Massachusetts Ave., NE
Washington, DC 20212
e-mail: zadrozny_p@bls.gov

July 5, 2005

Additional key words: Taylor-series approximation, model selection, numerical solution

JEL codes: C32, C43, C53, C63

[*] The paper's analysis and conclusions represent the authors' views and do not necessarily represent any positions of the Bureau of Economic Analysis or the Bureau of Labor Statistics.

ABSTRACT

For period t, let $q_t = f(v_t) + \tau_t$, where q_t denotes measured output quantity, $f(\cdot)$ denotes a production function, $v_t = (v_{1t}, \ldots, v_{nt})^T$ denotes a vector of n input quantities, τ_t denotes total factor productivity (TFP), and all variables are in natural-log form. Then, $f(v_t) = \sum_{i=1}^{n} \alpha_{it} v_{it}$, for $0 < \alpha_{it} < 1$ and $\sum_{i=1}^{n} \alpha_{it} = 1$, is a Cobb-Douglas (CD) 1st-order log-form approximation of a production function. If $f(\cdot)$ is approximated as a CD production function, the share parameters, α_{it}, are set to successive two-period-averaged observed input-cost shares, and the observed input quantities are considered optimal or input-cost minimizing, then, $q_t - \sum_{i=1}^{n} \alpha_{it} v_{it}$ is the log of Solow-residual TFP (STFP). STFP could be subject to positive or negative input-substitution bias for two reasons. First, the CD production function restricts all input substitutions to one. Second, observed inputs generally differ from optimal inputs, which respond to input-price- substitution effects. In this paper, we test the possible input-price-substitution bias of STFP in capital, labor, energy, materials, and services (KLEMS) inputs data for U.S. manufacturing from 1949 to 2001. (1) Based on maximum likelihood estimation, we determine a best 4th-order approximation of a CES-class production function. The CES class includes not only the standard constant elasticity of input substitution production functions but also includes so called tiered CES production functions, in which prespecified groups of inputs can have their own input-substitution elasticities and input-cost shares are parameterized (i) tightly as constants, (ii) moderately as smooth functions, or (iii) loosely as successive two-period averages. (2) Based on the best or optimal estimated production function, we compute the implied optimal TFP (OTFP) as $q_t - f(\hat{v}_t)$, where $f(\hat{v}_t)$ denotes the best estimated production function evaluated at the computed optimal inputs, $\hat{v}_t$. (3) For the data in percentage-growth (%Δ) form, we obtain two main conclusions: (i) relative to the average values of %ΔSTFP and %ΔOTFP of about 1%, |%ΔSTFP - %ΔOTFP| exceeds about 100% about 30% of the time in the first half of the sample period, so that %ΔSTFP is frequently significantly biased relative to %ΔOTFP; (ii) in the second half of the sample period, %ΔSTFP - %ΔOTFP is mostly close to its average of about .1%, so that %ΔSTFP is not significantly biased relative to %ΔOTFP.

1. Introduction.

The paper is specifically motivated as discussed in the preceeding abstract, but is also more generally motivated by the desire to accurately compute price indexes based on explicit forms of the functions being maximized. There are two main, mathematically identical, but economically different applications: computing price indexes of production inputs based on maximizing output of a production function for given input costs, as here, and computing price indexes of consumer goods based on maximizing utility of consumed goods for given expenditures, as in Zadrozny and Chen (2005). Here, we consider standard constant elasticity of input substitution (CES) production functions, with one input-substitution elasticity for all inputs, and more general tiered CES (TCES) production functions, with a different input-substitution elasticity for each prespecified group of inputs.

We are also interested in using even more general production functions, which we call generalized CES (GCES) production functions (see equation (6.1) below), in which each input can have its own price elasticity parameter, but, for brevity, limit the present applications to CES and TCES production functions. CES and TCES production functions have analytical solutions of their optimization problems. GCES production functions generally do not have analytical solutions except in special homothetic cases, such as the CES and TCES cases. Generally, optimization problems based on GCES production functions can be solved only numerically. In Zadrozny and Chen (2005), we describe in more detail than here the multi-step perturbation (MSP) method as a quick and accurate method for numerically solving the corresponding utility maximization problem. The MSP method, as the name implies, is a multi-step extension of the single-step perturbation method (Chen and Zadrozny, 2003).

Here, we could have used analytical CES and TCES solutions, but, for two reasons, use numerical solutions produced by the MSP method. First, we use the MSP method in order to test its accuracy in solving the static optimization problems. In all cases, we obtained nearly double-precision or about 14-decimal-digit accuracy when we checked the numerical MSP solutions against the analytical solutions, which encourages us to work in the future only with numerical solutions of GCES production functions. Second, we are interested in studying TFP bias by generalizing the Cobb-Douglas (CD) production function by adding nonlinear log-form Taylor-series terms up to a specified order. However, to do this tractably we must restrict the number of estimated parameters and we

do this by parameterizing the higher-order Taylor terms in terms of these CES-class production functions.

We proceed here entirely in log form for four reasons: (i) TFP and related price and quantity indexes are usually considered in log form; (ii) log-form variables are unit free, scaled equivalently, and, hence, lie mostly within or close to a unit sphere, which promotes numerical accuracy; (iii) log-form derivatives of the CES-class production functions are easier to derive, program, and compute with; and, (iv) comparisons with benchmark Solow residuals are easier in log form.

As noted, q denotes the log of the quantity of observed goods and services, $f(\cdot)$ denotes the log of output produced by the production function, and, $\hat{\tau}_t = q - \hat{q}$ denotes the log of the level of technology or TFP of $f(\cdot)$, where $f(\hat{v})$ more specifically denotes the log of optimal output produced by optimal log-form inputs, $\hat{v}$. Henceforth, $\hat{\tau}_t$, with the hat, denotes optimal TFP (OTFP) based on the best production-function model, where "best" is explained in section 3, and τ_t, without the hat, denotes Solow-residual TFP (STFP) based on the not necessarily best model and observed but not necessarily optimal inputs. To distinguish between q and $f(\cdot)$, we, respectively, refer to them as "goods and services" and "output." Let $p = (p_1, \ldots, p_n)^T$ denote an n×1 vector of logs of observed or computed input prices (superscript T denotes vector or matrix transposition) and let $v = (v_1, \ldots, v_n)^T$ denote an n×1 vector of logs of observed or computed input quantities. The context of whether input quantities or prices are observed or optimal-computed will be spelled out in each case. Whether prices are in nominal or real (deflated) units makes no difference, so long as real prices in a period are obtained by deflating each nominal price by the same value.

We assume $f(\cdot)$ is analytical, hence, for a sufficiently large k is arbitrarily well approximated by a kth-order Taylor series. Let $e(x) = (\exp(x_1), \ldots, \exp(x_n))^T$ for any n×1 vector $x = (x_1, \ldots, x_n)^T$. We write the input-cost line as $e(p)^T e(\hat{v}) = e(p)^T e(v)$, where p and v are given, so that $e(p)^T e(v)$ denotes observed expenditures on inputs and optimal $\hat{v}$ is computed. We consider the following output maximization problem: for given $f(\cdot)$, p, and v, maximize $f(\hat{v})$ with respect to $\hat{v}$, subject to $e(p)^T e(\hat{v}) = e(p)^T e(v)$. Because $\hat{\tau}_t$ is absent from the statement of the problem, it plays no role in its solution. Like Solow, we compute $\hat{\tau}_t$ residually: first $\hat{v}$, then $\hat{\tau}_t$. The difference with Solow is that $\hat{v}$ is computed as optimal and is not equated with observed v.

We consider only interior solutions which satisfy the usual 1st- and 2nd-order conditions (2.1), (2.2), and (2.5). As functional forms, we consider standard CES production functions (Arrow et al., 1961) and more general TCES production functions, which are multi-level generalizations of standard single-level one-input-elasticity CES functions (Sato, 1967; Burnside et al., 1995), that allow different input groups to have different substitution elasticities. For each production function, we solve for optimal inputs using the MSP method. In the CES and TCES cases (such that the CD case is a subcase of the CES case), we use analytical solutions to check the MSP method's accuracy, and, given the present successful application of the MSP method with CES and TCES production functions, in the future, we shall consider more general GCES production functions which do not have analytical solutions.

By a model we mean (i) a multiple-times-differentiable production function, $f(\cdot)$, (ii) a parameterization of $f(\cdot)$ over a data sample, and (iii) values of constant structural parameters which determine $f(\cdot)$ in the sample. We now consider three parameterizations in more detail: (a) unrestricted time-varying reduced-form parameters set every period to different values of the structural parameters; (b) time-varying reduced-form parameters restricted by a smooth function of constant structural parameters; and, (c) constant reduced-form parameters equal to constant structural parameters.

For example, $f(v_t) = \sum_{i=1}^{n} \alpha_{it} v_{it}$ denotes a period-t log-form CD production function for mean-adjusted data, whose reduced-form parameters, α_{it}, depend on constant structural parameters in the vector θ. In the typical case (a) of a data-producing agency, reduced-form parameters are unrestricted, are set year-to-year to relative input costs, and are statistically unreliable (have infinite estimated standard errors), because the number of estimated structural parameters, $\dim(\theta)$, equals the number of observations, nT: $\alpha_{it} = \theta_{it}$, for i = 1, ..., n and t = 1, ..., T, so that $\dim(\theta) = nT$. In the typical academic case (c) of an econometric analysis, the reduced-form parameters are constant over a sample in terms of structural parameters and are statistically reliable (but, perhaps, are not the best estimates because the reduced-form parameters are constant over the sample) because there are fewer estimated structural parameters than observations: $\alpha_{it} = \theta_i$, so that $\dim(\theta) = n < nT$. In the application in section 3, we also consider the in-between case (b), in which nT reduced-form parameters vary smoothly according to an integrated moving-average (IMA) process (Gardner, 1985), such that $\dim(\theta) < nT$.

What difference does the extra generality of going beyond the CD production function make? Normally, empirical validity is measured by residual size. In this case, we have output residuals, $q - \hat{q}$, and input residuals, $v - \hat{v}$. However, because OTPF and output residuals are identical, judging TFP's empirical validity using sizes of output residuals makes no sense. For example, statistically ideal zero output residuals imply zero log-TFP. Thus, instead, we propose judging TFP's empirical validity using an information criterion (IC) based on input residuals. The many IC that have been proposed differ in their propensities for choosing models with particular numbers of parameters. For example, Akaike's IC (1973) often picks less parsimonious models (i.e., with more parameters), while Schwarz's IC (1978) often picks more parsimonious models.

As usual, for a given data sample, we consider a model's parameter estimates and derived quantities like TFP as statistically reliable when the parameter estimates and derived quantities have finite standard errors. This occurs if and only if the degrees of freedom of the parameter estimates are positive. Among the models being considered, the one which minimizes a chosen IC is considered the best or empirically-most-valid model. An IC test based on input residuals for choosing the best model for computing TFP has several advantages. First, the test's justification does not depend on the method for estimating parameters. Second, the test can compare nonnested models. Third, the test does not require data on produced goods and services, q, although, of course, these data are ultimately needed to compute TFP.

By setting input-cost-share parameters year-to-year to observed input cost shares, a Solow-residual analysis treats observed inputs as optimal, so that input residuals are exactly zero, degrees of freedom of estimated parameters are exhausted, and, strictly, the estimated parameters and implied TFP have no statistical reliability. By contrast, by testing with an IC based on input residuals, we can select the empirically-most-valid model among CES, TCES, and possibly other models, compute the best model's implied TFP, and compare it with benchmark Solow residuals. Along the way, we can check the MSP method's accuracy by comparing analytical and MSP-based numerical solutions in the CES and TCES cases which have analytical solutions. We illustrate these ideas using annual data on capital, labor, energy, materials, and services (KLEMS) inputs in manufacturing industries, from 1949 to 2001, released to the public by the Bureau of Labor Statistics. Thus, we provide a method for computing the empirically-most-valid TFP, potentially more valid than STFP. In other words, we check the

robustness of STFP to deviations of the production function from the CD approximation which underlies STFP.

The remainder of the paper is organized as follows. Section 2 discusses using the MSP method to compute optimal inputs and residual inputs. Section 3 discusses the econometric design for choosing the best model. Section 4 compares the present direct-production-function approach to an indirect-cost-function approach, usually based on the translog cost function, and in doing so defines what we mean by input-substitution bias. Section 5 applies the MSP method and the econometric design to the KLEMS data: (1) it applies the MSP method to the KLEMS data to compute input residuals for the CES and TCES models being considered; (2) it selects the best model which minimizes the information criteria of Akaike (1973), Schwarz (1978), and Hurvich and Tsai (1989); and, (3) it computes OTFP based on the best model, computes the benchmark STFP, and compares the two TFP computations. Section 6 sums up the paper.

2. Using the MSP Method to Compute Optimal Inputs.

We now describe the MSP method for computing a 1st-order approximation ($k = 1$) of the optimal input function for an output maximizing problem based on any twice differentiable production function and any number of steps ($h \geq 1$). The extension of this MSP computation to a 4th-order approximation and any number of steps is discussed in detail in Zadrozny and Chen (2005). Although, we discuss only the 1st-order approximation here in order to minimize the technical details, in the application in section 5 we use the 4th-order approximation. The MSP method is a variant of the 1st-order method for computing a cost-of-living index described by Vartia (1983).

We exploit the property for simplifying computations that maximizing a function in a constraint set results in a solution equivalent to the one obtained by maximizing a monotonic transformation of the function in the same constraint set. In original units of measurement denoted by upper-case letters, the output maximization problem is: for given $F(\cdot)$, P, and V, maximize $F(\hat{V})$ with respect to $\hat{V}$, subject to $P^T\hat{V} = P^TV$, where $F(\cdot)$, P, V, and $\hat{V}$ denote antilogs of $f(\cdot)$ and the elements of p, v, and $\hat{v}$. Although the original-unit and log-unit formulations of the problem lead to slightly different 1st-order conditions, they have equivalent solutions, namely, $\hat{V} = \exp(\hat{v})$. As noted before, proceeding in log form has several advantages.

We want to compute optimal and residual inputs, for each period, in a sample of input prices and quantities, for CES and TCES production functions. Let $\{p_t, v_t\}_{t=1}^{T}$ denote a given sample of observed input prices and quantities. Then, for given $f(\cdot)$, p_t, and v_t in period t, the vector of optimal inputs, $\hat{v}_t$, which solves the output maximization problem, implies the vector of input residuals, $v_t - \hat{v}_t$.

Figure 1 illustrates MSP computation of $\hat{v}_t$ in terms of two inputs, in the movement from points A to B. Points A and B denote start and end points of an MSP computation. Straight lines AA and BB, through A and B, denote start and end input-cost lines. Curved lines f_A and f_B, tangent at A to AA and tangent at B to BB, denote start and end isoquants. Observed input prices and quantities are $p = (p_1, p_2)^T$ and $v = (v_1, v_2)^T$. Observed v is at A and BB denotes the "observed" cost line defined by observed p and v: $e(p)^T e(\hat{v}) = e(p)^T e(v)$. The objective is to compute $\hat{v}$, the optimal combination of inputs on BB. The implied negative input residual, $\hat{v} - v$, is depicted by the vector difference B - A.

The MSP method starts at A but generally works correctly only if the starting point is optimal. Generally, A is not optimal on the observed cost line BB, because isoquant f_A, which passes through A, is not tangent to BB at A. However, A is optimal on AA, because AA is constructed to be tangent to f_A at A. Accordingly, AA is defined by $e(\hat{p})^T e(\hat{v}) = e(\hat{p})^T e(v)$, where $\hat{p}$ satisfies the 1st-order conditions (2.1) and (2.2), for given $f(\cdot)$ and v. Thus, $\hat{p}$ and AA are "optimal" at A. The MSP method computes the change in optimal inputs as they move from A to B in response to the counterclockwise rotation of the input-cost line at the initial point A, as the price vector flattens from $\hat{p}$ in AA to p in BB.

As before, for given assumed $f(\cdot)$ and given observed p and v, the objective is to compute optimal $\hat{v}$. For these given quantities, the log-form output-maximization problem is: maximize $f(\hat{v})$ with respect to $\hat{v}$, subject to $e(p)^T e(\hat{v}) = e(p)^T e(v)$. The Lagrangian function of the problem is $\ell = f(\hat{v}) + \hat{\lambda}(e(p)^T(e(v) - e(p)^T e(\hat{v}))$, where $\hat{\lambda}$ denotes the Lagrange multiplier. We obtain the 1st-order conditions of the maximization problem by differentiating ℓ with respect to $\hat{v}$ and $\hat{\lambda}$, set the results to zero, and write them as

$$\nabla f(\hat{v}) = \hat{\lambda} e(p + \hat{v})^T, \tag{2.1}$$

(2.2) $\quad e(p)^{T}e(\hat{v}) = e(p)^{T}e(v),$

where $\nabla f(\hat{v}) = [\partial f(\hat{v})/\partial v_1, \ldots, \partial f(\hat{v})/\partial v_n]$ denotes the 1×n gradient row vector of first-partial derivatives of $f(\hat{v})$. For given $f(\cdot)$, p and v, equations (2.1) and (2.2) can be solved for unique values of $\hat{v}$ and $\hat{\lambda}$, at least locally and numerically if 2nd-order conditions (2.5) hold.

As discussed before, we start the MSP method at observed inputs and need to treat them as optimal. Because observed inputs, v, are generally not optimal at observed prices, p, we first need to compute the "optimal" price vector, $\hat{p}$, at which v is optimal. We do this by considering the 1st-order conditions (2.1) and (2.2) as $\nabla f(v) = \hat{\lambda} e(\hat{p}+v)^{T}$ and $e(\hat{p})^{T}e(v) = e(p)^{T}e(v)$, for given assumed $f(\cdot)$ and given observed p and v, and solving for $\hat{\lambda}$ and $\hat{p}$. Let $E(x) = \mathrm{diag}(e(x))$ denote the n×n diagonal matrix with n×1 vector e(x) on the principal diagonal; because all observed inputs are positive and finite, E(v) has finite and nonzero diagonal elements and, hence, is nonsingular; $E(v)^{-1}e(v) = u$, where $u = (1, \ldots, 1)^{T}$ denotes the n×1 unit vector of ones; and, $e(\hat{p})^{T}e(v) = e(p)^{T}e(v)$ when computing $\hat{p}$, because the computed input-cost line defined by $\hat{p}$ and the observed input-cost line defined by p both pass through the observed inputs, v. The solution values of $\hat{\lambda}$ and $\hat{p}$ are

(2.3) $\quad \hat{\lambda} = \nabla f(v) u / e(p)^{T}e(v),$

$$e(\hat{p}) = E(v)^{-1}\nabla f(v)^{T}/\hat{\lambda}.$$

At this point, having computed $\hat{p}$ according to equations (2.3), we now consider $\hat{p}$ as observed and given, and relabel it as p. Thus, we now consider as given the same $f(\cdot)$ and v as before and the computed $\hat{p}$ relabelled as p. For these given quantities, we now differentiate 1st-order conditions (2.1) and (2.2) with respect to $\hat{v}$, $\hat{\lambda}$, and p and write the result as

(2.4) $\quad F(x)d\hat{y} = G(x)dp,$

$$\text{or} \qquad \begin{bmatrix} \nabla^2 f(\hat{v}) - \hat{\lambda} E(p + \hat{v}) & -e(p + \hat{v}) \\ -e(p + \hat{v})^T & 0_{1\times 1} \end{bmatrix} \begin{bmatrix} d\hat{v} \\ d\hat{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{\lambda} E(p + \hat{v}) \\ e(p + \hat{v})^T - e(p + v)^T \end{bmatrix} dp,$$

where $\nabla^2 f(\cdot)$ denotes the n×n Hessian matrix of 2nd-partial derivatives of $f(\cdot)$, $F(x)$ is an (n+1)×(n+1) matrix function, $G(x)$ is an (n+1)×n matrix function, $x = (\hat{y}^T, p^T)^T$ contains all 2n+1 variables, $\hat{y} = (\hat{v}^T, \hat{\lambda})^T$ contains the n+1 "endogenous" variables to be determined, and p contains the n given or "exogenous" prices. Although the hats emphasize computed values, for simplicity, we omit them from x because, unlike in v or y, we do not need to distinguish between hatted and unhatted x.

The elements of x are all known, because they are either observed or previously computed. For given x, equation (2.4) implies the unique value $d\hat{y} = H(x)dp$, where $H(x) = F(x)^{-1}G(x)$, if and only if $|F(x)| \neq 0$, which is implied by the 2nd-order conditions of the problem, where $|\cdot|$ denotes the determinant of a square matrix. For $i = 2, \ldots, n$, let $F_i(x) = \begin{bmatrix} A_i(x) & b_i(x) \\ b(x)^T & 0_{1\times 1} \end{bmatrix}$ denote the (i+1)×(i+1) submatrix of $F(x)$, where $A_i(x)$ denotes the upper-and-left-most i×i submatrix of $\nabla^2 f(\hat{v}) - \hat{\lambda} E(p+\hat{v})$ and $b_i(x)$ denotes the first i elements of $-e(p+\hat{v})$. Then, the 2nd-order conditions of the output-maximization problem are

$$(2.5) \qquad (-1)^{i+1}|F_i(x)| > 0,$$

for $i = 2, \ldots, n$ (Mann, 1943; Samuelson, 1947). Thus, when x maximizes output and satisfies 2nd-order condition (2.5), equation (2.4) has the unique solution

$$(2.6) \qquad d\hat{y} = H(x)dp,$$

where $H(x) = F(x)^{-1}G(x)$ is an (n+1)×n matrix function of x. Although equation (2.6) derives from the true y process, we write its left side as $d\hat{y}$ to emphasize that the true dy is approximated using this equation.

We now consider an interaction between continuous and discrete time. Let $[1,T+1) = \bigcup_{t=1}^{T} [t, t+1)$ denote a continuous-time interval divided into T unit-length periods indexed by their beginning moments, $t = 1, \ldots, T$, where $[t,t+1) = \{s | t \leq s < t+1\}$. Definitions of variables hold both in continuous time within

a period, denoted by argument s, and in discrete time t at starting moments of the periods, denoted by subscript t. Thus, discrete time periods are indexed by their starting moments. Above, we denoted observed and given values without hats and computed values with hats. We now also denote true values without hats and continue to denote computed values with hats. For example, y(s) denotes true y in continuous time and $\hat{y}(s)$ denotes computed y in continuous time. Because computed values are meant to be optimal but are actually approximations of optimal values, strictly, a hat implies a value is "computed, optimal, and approximate," although for simplicity, we refer to hatted values only as computed.

For each period t, an MSP computation proceeds as follows. We think of starting computations at the start of a period, at the moment s = t, and ending them at the end of the period, at the moment s = t+1. We think of the observed input quantities, v_t, as occuring at the start of the period and think of the observed input prices, p_t, as occuring at the end of the period. For given assumed $f(\cdot)$ and given observed p_t and v_t, we first compute the "optimal" starting price vector, $\hat{p}_t$, which makes v_t optimal. We assume the price vector moves continuously from its "optimal" starting value, $\hat{p}_t$, to its observed ending value, p_t. Then, given $\hat{v}(t_1) = v_t$, we compute the remaining optimal input quantities at h equidistant points along the optimal input path in response to the price movements. We compute $\Delta\hat{y}_{t_i} \equiv \hat{y}(t_{i+1}) - \hat{y}_{t_i} \equiv \int_{s=t_i}^{t_{i+1}} d\hat{y}(s)$, for i = 1, ..., h, and pick $\Delta\hat{v}_t \equiv \hat{v}(t_h) - v_t$ as the top n×1 subvector of $\Delta\hat{y}_t = \sum_{i=1}^{h} \Delta\hat{y}_{t_i}$, so that $\hat{v}_t = v_t + \Delta\hat{v}_t$. Figure 1 depicts the computations as the movement from points A to B along the curved line with arrowheads.

The implicit function theorem (Apostol, 1974, p. 374), upon which the MSP method is based, implies that if the production function is twice differentiable and satisfies the 2nd-order conditions, so that its optima are interior points, then, the exact solution path is differentiable, and, in each computational subperiod $s \in [t_i, t_{i+1})$, for i = 1, ..., h, has the 1st-order polynomial Taylor-series approximation of the true y(s),

$$\hat{y}(s) = \hat{y}_{t_i} + \nabla\hat{y}_{t_i}(s-t_i), \tag{2.7}$$

where $\nabla\hat{y}_{t_i}$ is an (n+1)×1 coefficient to be computed in terms of observed input prices and quantities. We state an analogous polynomial price process in equation (2.10).

The approximation $\Delta\hat{y}_t = \sum_{i=1}^{h} \Delta\hat{y}_{t_i}$ of $\Delta y_t \equiv y(t+1) - y_t \equiv \int_{s=t}^{t+1} dy(s)$ has the theoretical approximation error $\varepsilon = |\Delta y_t - \Delta\hat{y}_t|$. We partition each period $[t,t+1)$ into h subperiods of length h^{-1}, as $[t,t+1) = \bigcup_{i=1}^{h} [t_i, t_i+h^{-1})$, where $[t_i,t_i+h^{-1}) = [t+(i-1)h^{-1}, t+ih^{-1})$, for i = 1, ..., h. For each subperiod i in period t, we compute the coefficient of the approximate process (2.7), $\nabla\hat{y}_{t_i}$, and, then, compute the subperiod increments, $\Delta\hat{y}_{t_i}$, as

$$\Delta\hat{y}_{t_i} = \nabla\hat{y}_{t_i} h^{-1}. \tag{2.8}$$

The theoretical approximation error of a kth-order approximate solution is on the order of h^{-k}. The approximation error can be controlled by setting k and h, although, in the discussion in this section, we consider only h > 1 and k = 1. See Zadrozny and Chen (2005, table 1) for values of h and k which predict achieving particular orders of magnitude of accuracy.

We now describe the MSP method for k = 1. For i = 1, ..., h and $s \in [t_i,t_i+h)$, we differentiate the approximate y process (2.7) with respect to s and obtain

$$d\hat{y}(s) = \nabla\hat{y}_{t_i}. \tag{2.9}$$

We compute the coefficient, $\nabla\hat{y}_{t_i}$, so that it is equal to the first differential of the true y process (2.6).

Analogous to approximate y process (2.7), we assume prices follow a 1st-order polynomial process, for $s \in [t,t+1)$ and t = 1, ..., T,

$$p(s) = \hat{p}_t + \nabla p_t (s-t), \tag{2.10}$$

with n×1 coefficient ∇p_t. Although the price coefficient remains at its initial value, indexed at $t_1 = t$, throughout computations in period t, the y coefficient, $\nabla\hat{y}_{t_i}$, is indexed by t_i and updated at each iteration i within period t. From price process (2.10), we require only that it passes through the

computed start-of-period prices $p(t) = \hat{p}_t$, given by equations (2.3), and the observed end-of-period prices $p(t+1) = p_t$, because in discrete time firms care only about starting and ending prices and do not care about within-period prices.

In the following, we distinguish between differentiating with respect to s and differencing with respect to t. For $s \in [t,t+1)$, differentiating price process (2.10) with respect to s, we obtain

$$dp(s) = \nabla p_t. \tag{2.11}$$

Then, differencing price process (2.10), we obtain the price coefficient, ∇p_t, in terms of differenced prices, Δp_t, as

$$\nabla p_t = \Delta p_t. \tag{2.12}$$

where the differenced prices are given in terms of the computed and observed prices, $\hat{p}_t$ and p_t, as

$$\Delta p_t = p_t - \hat{p}_t. \tag{2.13}$$

As required, the price coefficient set by equations (2.12) and (2.13) implies that the price process (2.10) passes through the computed and observed prices.

We now describe computing $\{\hat{v}_t\}_{t=1}^{T}$ using the MSP method. We sequence the computations in an outer loop, for periods t = 1, ..., T, and an inner loop, for subperiods i = 1, ..., h. For each period t, we describe the inner-loop computations in four steps. Within the four steps, we take as given an assumed $f(\cdot)$ and observed p_t and v_t.

Step 1: Initialize x, Prices, and Their Differentials.

For given $f(\cdot)$, p_t, and v_t and for i = 1, hence, for $s = t_1 = t$, we first compute $\hat{\lambda}_t$ and $\hat{p}_t$ according to equations (2.3). We set $x_t = (\hat{y}_t^T, \hat{p}_t^T)^T = (v_t^T, \hat{\lambda}_t, \hat{p}_t^T)^T$. Following equation (2.11), we set the price differential as $dp(t) = \nabla p_t$. Following equations (2.12) and (2.13), we compute the price coefficient, ∇p_t, in terms of the computed and observed prices, $\hat{p}_t$ and p_t.

Step 2: Compute 1st-Order y Coefficient.

For $s = t_1 = t$, equations (2.6), (2.9), and (2.11) imply that

(2.14) $H(x_t) = F(x_t)^{-1}G(x_t)$,

$\nabla\hat{y}_t = H(x_t)\nabla p_t$.

For $k = 1$ and $i = 1$, equation (2.8) implies that

(2.15) $\Delta\hat{y}_t = \nabla\hat{y}_t h^{-1}$,

so that $\hat{y}_{t_2} = \hat{y}_t + \Delta\hat{y}_t$.

Step 3: Update Prices, x, and y.

For $i = 2$ and, hence, for $s = t_2 = t+h^{-1}$, equations (2.10) and (2.11) imply that we update prices and their differentials as

(2.16) $p(t_2) = \hat{p}_t + \nabla p_t h^{-1}$,

$dp(t_2) = \nabla p_t$,

such that the price coefficient, ∇p_t, remains at its initial $t_1 = t$ computed value. We set $x_{t_2} = (\hat{y}_{t_2}^T, p_{t_2}^T)^T$. We repeat step 2 and, thereby, update the y coefficient to $\nabla\hat{y}_{t_2}$. Following equation (2.8), we compute $\Delta\hat{y}_{t_2} = \nabla\hat{y}_{t_2} h^{-1}$ and $\hat{y}_{t_3} = \hat{y}_{t_2} + \Delta\hat{y}_{t_2}$.

Step 4: Repeat Steps 2 and 3.

For $i = 3$ and, hence, for $s = t_3 = t+2h^{-1}$, we update prices and their differentials as

(2.17) $p(t_3) = \hat{p}_t + 2\nabla p_t h^{-1}$,

$$dp(t_3) = \nabla p_t .$$

We set $x_{t_3} = (\hat{y}_{t_3}^T, p_{t_3}^T)^T$. We repeat step 2 and update the y coefficient to $\nabla\hat{y}_{t_3}$. We compute $\Delta\hat{y}_{t_3} = \nabla\hat{y}_{t_3} h^{-1}$ and update y as $\hat{y}_{t_4} = \hat{y}_{t_3} + \Delta\hat{y}_{t_3}$. We repeat these steps for i = 4, ..., h and, hence, for $s = t_4 = t+3h^{-1}$, ..., $t_h = t+1-h^{-1}$. Finally, we compute $\Delta\hat{y}_{t_h}$ and pick $\Delta\hat{v}_{t_h}$ as the top n-dimensional subvector of the computed $\Delta\hat{y}_{t_h}$.

3. Econometric Design.

We now discuss the econometric design of the empirical application. As noted before, a model is a production function, a parameterization of the production function over a sample, and particular numerical values of the constant structural parameters in the vector θ. The structural parameters could determine time-varying processes of parameters more directly in the production function. For example, in the IMA models in section 5, production-function share parameters follow integrated moving averages (IMA) defined by elements of θ. The ultimate goal is maximum likelihood estimation of several models and choosing as the best one the model which minimizes one or more information criteria (IC). However, because we do not yet have all the necessary computer programs completed, for now we apply a coarse version of maximum likelihood estimation. That is, for each considered class of models (CES and TCES), we pick a best model from a set of models defined over a relatively small and discrete grid of numerical parameter values. When the additional computer programming is done, we shall be able to implement the maximum likelihood estimation more fully over continuous intervals of the parameters.

For a particular model, the log-likelihood function and an IC are computed as follows. Suppose we have a sample of observations on input prices and quantities, in log form $\{p_t, v_t\}_{t=1}^T$, for periods t = 1, ..., T. Period-t log-form input residuals are observed input quantities minus computed optimal input quantities, $\xi_t = v_t - \hat{v}_t$. Suppose the residuals are distributed normally, identically, independently, with zero means, and covariance matrix Σ_ξ or $\xi_t \sim$ NIID(0, Σ_ξ). Let $L(\theta)$ denote -(2/T) × log-likelihood function, except for terms independent of parameters. Then, $L(\theta) = \ln|\hat{\Sigma}_\xi|$, where $\ln|\cdot|$ denotes the natural

logarithm of a determinant and $\hat{\Sigma}_\xi = (1/T)\sum_{t=1}^{T}\xi_t\xi_t^T$, where the residuals, ξ_t, are evaluated at a particular values of θ. An IC = $L(\theta)$ + $P(\dim(\theta))$, where $P(\dim(\theta))$ denotes a penalty term which depends on the number of estimated parameters, $\dim(\theta)$. Each structural parameter is estimated, so that $\dim(\theta)$ = number of structural parameters. In section 5, we consider Akaike's (1973) information criterion (AIC), Hurvich and Tsai's (1989) bias-corrected AIC (BCAIC), and Schwarz's (1978) Bayesian information criterion (BIC). For example, for AIC, $P(\dim(\theta)) = (2/T)\dim(\theta)$. In each model class, we choose as the maximum likelihood estimate the model which minimizes $L(\theta)$ in the class over the parameter grid. The theory behind ICs says that they can choose a best model among nested or nonnested models. As the best overall model, we choose the one which minimizes one or more ICs.

4. Comparing with Translog Cost Function and Defining Substitution Bias.

We now discuss the advantages of the present direct production-function approach compared with an indirect cost-function approach, in particular, compared with the translog cost-function approach, the most commonly used indirect cost function. In doing so, we define "substitution bias." Notation is as before; in particular, lower-case letters denote logarithms.

The direct output-maximization problem is: for a given production function and observed input prices and quantities, $f(\cdot)$, p, and v, maximize output, $f(\hat{v})$, with respect to input quantities, $\hat{v}$, subject to the cost line, $e(p)^T e(\hat{v}) = e(p)^T e(v)$. Under 1st- and 2nd-order conditions, the problem has a unique solution for given p and v, $\hat{v} = g(p,v)$. The application in section 5 is based on a purely numerical 4th-order approximation of $g(p,\cdot)$, for varying p and constant v,

$$\begin{aligned}(4.1)\qquad \hat{v} = g(p,\cdot) \cong v + \nabla g(p - p') + (1/2)[(p - p')^T \otimes I_n]\nabla^2 g(p - p') \\ + (1/6)[(\Pi_2 \otimes (p - p')^T \otimes I_n]\nabla^3 g(p - p') \\ + (1/24)[(\Pi_3 \otimes (p - p')^T \otimes I_n]\nabla^4 g(p - p'),\end{aligned}$$

where p′ denotes the computed initial input-price vector in a period, as in section 2, ∇g, ..., $\nabla^4 g$ denote matrices of 1st- to 4th-order partial

derivatives of g with respect to p evaluated at p' and v, $\Pi_k \otimes (p-p')^T$ denotes k-1 successive Kronecker products of $(p-p')^T$, $\otimes$ denotes a single Kronecker product, I_n denotes the n×n identity matrix, and n = dim(p). Chen and Zadrozny (2003, appendix A) discuss this notation in more detail.

The corresponding indirect cost-minimization problem is: for given $f(\cdot)$, p, and $q' = q-\tau$, minimize input costs, $c = \ln[e(p)^T e(\hat{v})]$, with respect to input quantities, $\hat{v}$, subject to the production function, $f(\hat{v}) = q'$. Under 1st- and 2nd-order conditions, the problem has a unique solution, $\hat{v} = h(p,q')$, so that the minimized indirect cost function is a function of p and q', $\hat{c}(p,q') = \ln[e(p)^T e(h(p,q'))]$. The indirect-cost-function approach was introduced to circumvent the inability to solve analytically (i.e., explicitly and in closed form) direct problems based on more general production functions than the CES production function, which was considered empirically too limited (see also Berndt, 1991, ch. 9, pp. 449-506).

The most frequently used indirect cost function is the translog cost function, a 2nd-order Taylor-series approximation of $\hat{c}(p,q')$ (Christensen et al., 1971, 1973). Like equation (4.1), the translog approximation of $\hat{c}(p,q')$ only in terms of prices is

$$(4.2) \qquad \hat{c}(p,\cdot) \cong \hat{c}(p',q') + \nabla\hat{c}(p - p') + (1/2)(p - p')^T \nabla^2\hat{c}(p - p'),$$

where $\nabla\hat{c}$ and $\nabla^2\hat{c}$ are 1×n and n×n matrices of 1st- and 2nd-partial derivatives of $\hat{c}(p,q')$ with respect to p evaluated at p' and v. The envelope theorem (also called Shepard's lemma and Roy's theorem) implies that the 1st-partial derivatives of $\hat{c}(p,\cdot)$ with respect to p are equal to the optimal input function. Thus, differentiating equation (4.2) with respect to p and using the symmetry of $\nabla^2\hat{c}$ implies that

$$(4.3) \qquad \hat{v} = h(p,\cdot) \cong \nabla\hat{c}^T + \nabla^2\hat{c}(p - p'),$$

a 1st-order approximation of the optimal input function, so that equation (4.3) corresponds to the first two terms on the right side of equation (4.1). Thus, the 4th-order approximate optimal input function used here is more general, at least in certain dimensions, than the 1st-order function (4.3).

We now define substitution bias. Taylor-series theory says that errors in approximations (4.1) and (4.3) of the optimal input function are, respectively, on the order of $||p-p'||^5$ and $||p-p'||^2$, where $||\cdot||$ denotes a vector norm (Golub and Van Loan, 1996, pp. 52-54). Because optimal inputs are known to be homogeneous of degree zero in p and c, the approximation errors should be a concern only when individual prices change in different proportions. In any period, the difference between equations (4.1) and (4.3) is

$$\delta = (1/6)[(\Pi_2 \otimes (p-p')^T \otimes I_n]\nabla^3 g(p-p') + (1/24)[(\Pi_3 \otimes (p-p')^T \otimes I_n]\nabla^4 g(p-p'). \tag{4.4}$$

The quantity δ assumes positive and negative values over a sample of periods. We say significant input-substitution bias exists in a sample if δ has a significant nonzero mean, a significant variance, or both, where "significance" could be interpreted according to the subject matter of the application. Equation (4.4) implies input-substitution bias occurs if and only if $||p-p'||$ is sufficiently large. Significant input-substitution bias carries over through the production function to TFP. In section 5, we conclude that STFP, based on a 1st-order production-function approximation, is frequently significantly biased relative to OTFP based on a 4th-order production-function approximation.

The direct approach is preferred because it is easier to use for parsimoniously generalizing a model. In empirical work, we want a model to be general, so that it fits data well, but also want it to be parsimonious, so that the estimated parameters, the estimated model, and any quantities derived therefrom are statistically significant. Thus, we seek a balance between model generality and parsimony. The direct approach is easier to use for this purpose. For example, section 5 illustrates successful parsimonious generalization of the standard constant-parameter CES model to TCES models with time-varying parameters. By contrast, although cost function (4.2) is straightforwardly extended to the 4th order, doing so effectively is not easy. The extension adds many new unrestricted coefficients in the derivative matrices $\nabla^3 \hat{c}$ and $\nabla^4 \hat{c}$, even after imposing homogeneity restrictions, which cannot effectively all be treated as new parameters to be estimated; the new coefficients must somehow be parameterized more tightly. Moreover, curvature restrictions implied by the 2nd-order conditions must also be maintained. It is not clear how this should be done. Generalizing a model using the combined direct and MSP methods seems to be easier.

5. Application to KLEMS Data.

We now discuss the application to annual data for U.S. manufacturing from 1949 to 2001 from the Bureau of Labor Statistics (2002). The data are prices and quantities of capital (K), labor (L), energy (E), materials (M), and services (S) used by U.S. manufacturing firms to produce output. The raw data are indexes of input quantities (with 1996 values being 100), expenditures on inputs in billions of current dollars, and the value of output in billions of current dollars. Prices of inputs are computed as expenditures divided by input quantity indexes. As noted before, the scale of prices makes no difference, in particular, whether they are in current- or constant-dollar form.

STFP is based on a 1st-order CD approximation of any differentiable production function. Here, a production function parameterized in a certain way is a model. We consider CES and TCES models of the five KLEMS inputs, such that in unit-elastic cases a CES model reduces to a CD model. The parameters are input-cost shares, α_1, ..., α_5, and input substitution elasticities, σ_1 in the CES models and σ_1 and σ_2, for $\sigma_1 > \sigma_2$, in the TCES models. For the 53 years, we consider "constant" α_i's estimated as sample means, "IMA" α_i's equal to one-period ahead forecasts of estimated IMA(1,1) models of the cost shares, and "Törnqvist" α_i's set to .5 × period t's observed input-cost shares + .5 × period t-1's observed input-cost shares. We estimate the IMA parameters by applying maximum likelihood estimation (MLE) to the observed cost shares. In each case, because the cost shares must sum to one, we set the α_i's of the four largest LMKS-cost shares and set the remaining E-cost shares residually, as one minus the sum of the other α_i's. For both CES and TCES models, we consider σ_1 and $\sigma_2 \in \{.1, .18, .5, .67, 1, 1.5\ 2, 5.9, 10\}$. Thus, we do a coarse MLE over a small and discrete grid of σ_1's, conditional on the estimated α's. We do not consider joint MLE of parameters, because this usually results in implausible estimates of α_i's. For example, until he introduces utilization rates (an extension which is beyond the scope of this paper), Tatom (1980) obtains the estimates $\alpha_L > 1$ and $\alpha_K < 0$, which contradict diminishing and positive marginal productivities of labor and capital.

We evaluate the estimated models in terms of AIC, BCAIC, and BIC. We are especially concerned about degrees of freedom (DF) of estimated parameters and, for a particular IC, consider as the best one the model which minimizes that IC for positive DF. We are concerned with DF because a model with zero DF implies that the model's estimated parameters and any quantities such as TFP derived

therefrom have infinite variances and, hence, strictly have no statistical reliability. To varying extents, the ICs considered here account for DF in their penalty terms. Among the ICs in table 1, BCAIC most effectively accounts for DF, because it is the only IC that approaches $+\infty$ as DF approach zero from above. Thus, we set BCAIC = $+\infty$ when DF are nonpositive. An IC is parsimonious if it selects as the best one the model with the fewest parameters. The ICs in tables 1 are ordered in increasing parsimony according to AIC, BCAIC, and BIC.

5.1. Results from CES-Class Models.

We considered nine classes of models determined by three classes of production functions (CES, $TCES_1$, $TCES_2$) and three input-cost-share parameterizations (constant, IMA, Törnqvist). For each model class, table 1 reports MLEs of σ_i over the discrete parameter grid, DF, $-(2/T)L(\theta)$, AIC, BCAIC, and BIC. In the cases of $\sigma_1 = 1$, CES models reduce to CD models.

The DF in table 1 are obtained as follows. Each model has five KLEMS inputs. Because cost shares sum to one, there are four free cost shares in each of the 53 years. Each model also has one or two elasticity parameters, σ_1 and σ_2. Thus, constant-cost-share models 1, 4, and 7 have 5 and 6 estimated parameters, hence, have 47 and 46 DF. Each IMA process has two estimated parameters, a moving-average coefficient and a white-noise disturbance variance. Thus, IMA-cost-share models 2, 5, and 8 have 9 and 10 estimated parameters, hence, have 43 and 42 DF. Finally, Törnqvist-cost-share models 3, 6, and 9 have 213 and 214 estimated parameters, hence, have zero DF. Figure 2 depicts the largest cost-share inputs, L, M, K, and S. That is, the smallest cost shares of E are not graphed. In figure 2, each panel contains time plots of constant, IMA, and Törnqvist cost shares for each of the LMKS inputs. Strictly each panel has three cases, but practically each panel has two cases, because the IMA and Törnqvist graphs are nearly identical. Thus, the IMA and Törnqvist models differ significantly only in their DF.

Summarizing the results in table 1 for the CES models: in the constant-cost share case, $\sigma_1 = .5$ yields the best minimal IC values; IMA-cost-share model 2 is the best CES model, because it has the lowest ICs for positive DF; and, although Törnqvist-cost-share model 3 (and models 6 and 9) has lower ICs than model 2, we consider it inferior because it has zero DF.

5.2. Results from TCES Models

Even if we limit the TCES model search to two-tiered models, this results in more models than we could evaluate in practice. Thus, we first looked at figures 3 and 4 to obtain guidance about which input groups to form. Figure 3 depicts the 10 pairwise scatter plots of the KLEMS inputs in log form. In the figure, all pairwise plots except those involving L follow clear, noiseless, mostly upward, straight or curved lines. Plots involving L are quite noisy. Thus, figure 3 suggests that all non-L inputs move in close to fixed proportions and have low substitutability. That is, figure 3 suggests a two-tiered TCES model with an outer group of L and KEMS, with relatively high input substitution σ_1, and an inner group of K, E, M, and S, with relatively low input substitution σ_2. Thus, we consider the L-KEMS two-tiered CES model, denoted $TCES_1$, written in original unlogged form as

$$(5.1) \qquad Q = [\alpha_1 L^{\rho} + \alpha_2(\beta_1 K^{\gamma} + \beta_2 E^{\gamma} + \beta_3 M^{\gamma} + \beta_4 S^{\gamma})^{\rho/\gamma}]^{1/\rho},$$

where α_i, $\beta_i > 0$, $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$, and $\gamma < \rho < 1$; the outer group, L and KEMS, has $\sigma_1 = (1 - \rho)^{-1}$ and, the inner group, K, E, M, and S, has $\sigma_2 = (1 - \gamma)^{-1}$, so that $\sigma_1 > \sigma_2$.

Figures 4a-b suggest a two-tiered TCES model with L-E-KMS input groups. Figure 4a depicts the following broad input-price movements: all input prices except E prices follow the same upward trend, exhibit relatively minor differences about the trend, and E prices are relatively constant from 1949 to 1972 and from 1982 to 2001 and rise sharply from 1973 to 1981. Figure 4b depicts the following broad input-quantity movements: L is relatively constant; K, M, and S follow each other very closely along an upward trend; and, E rises significantly until 1973 and thereafter grows very slowly. In particular, figure 4b suggests a two-tiered TCES model with an outer group of L, E, and KMS, with relatively high input substitution σ_1, and an inner group of K, M, and S, with relatively low input substitution σ_2. Because figure 4b shows that K, M, and S move in close to fixed proportions, we expect σ_2 to be relatively small. The relative constancy of L in figure 4b could also be interpreted as indicating nonneutral L-saving technical change, but we limit the analysis to homothetic production functions, hence, limit it to the neutral technical change of the STFP. Thus, we consider the L-E-KMS two-tiered CES model, denoted $TCES_2$, written in original unlogged form as

(5.2) $$Q = [\alpha_1 L^{\rho} + \alpha_2 E^{\rho} + \alpha_3(\beta_1 K^{\gamma} + \beta_2 M^{\gamma} + \beta_3 S^{\gamma})^{\rho/\gamma}]^{1/\rho},$$

where α_i, $\beta_i > 0$, $\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 1$, and $\gamma < \rho < 1$; the outer group, L, E, and KMS, has $\sigma_1 = (1 - \rho)^{-1}$ and, the inner group, K, M, and S, has $\sigma_2 = (1 - \gamma)^{-1}$, so that $\sigma_1 > \sigma_2$.

Table 1 reports MLEs of σ_1 and σ_2, DF, $-(2/T)L(\theta)$, AIC, BCAIC, and BIC for the TCES models. Because there are outer and inner elasticities of input substitution in the TCES models, DF equals 46 in the constant-cost share models, 42 in the IMA-cost share models, and remains zero in the Törnqvist-cost share models. In the $TCES_1$ models, IMA-cost-share model 5, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs for positive DF. Similarly, in the $TCES_2$ models, IMA-cost-share model 8, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs for positive DF.

Table 1 implies that $TCES_1$ IMA-cost-share model 5 is the best model among those being considered, because it has the lowest ICs for positive DF. Thus, table 1 rejects a single elasticity of input substitution for all KLEMS inputs. Table 1 also implies that the best model 5 dominates the CD Törnqvist-cost-share model underlying STFP. Model 3 in the table is also a CD Törnqvist-cost-share model, but differs from the STFP model because its ICs are based on optimal inputs, not on observed inputs, which generally differ from optimal inputs. Thus, the STFP model's ICs are greater (inferior; actually infinite) than those of models 3 and 5 and, consequently, STFP is statistically less appropriate than the OTFP of models 3 or 5. Strictly, model 3 also differs from the STFP model because its cost shares are IMA, not Törnqvist. However, because IMA and Törnqvist cost shares follow each other very closely, as figure 2 indicates, the input-cost-share differences between these models should not cause their ICs to differ much.

The MSP method was accurate for all models and sample periods. There are six 1st-order conditions (FOC), five marginal productivity conditions of the KLEMS inputs and the cost line. Ideally, each computed FOC is zero, but, in practice, the best we can do is to compute each FOC up to a small remainder, called the FOC residual. The overall accuracy of the computational method, whether MSP or any other method, can be measured by the largest absolute FOC residual. In the application, the MSP method computed absolute FOC residuals no larger than about 10^{-14}. Because the data contained no more than 6 decimal digits, computed FOC residuals no larger than about 10^{-14} represent very accurate computations.

5.3. Optimal TFP Compared with Solow-Residual TFP.

We used the best $TCES_1$ IMA-cost-share model 5 to compute year-to-year percentage growth in OTFP or $\%\Delta OTFP = \Delta q - \Delta f(\hat{v})$, where Δq and $\Delta f(\hat{v})$ denote year-to-year percentage growth in observed output and computed optimal output based on model 5. Similarly, $\%\Delta STFP = \Delta q - \alpha_k \Delta k - \ldots - \alpha_S \Delta s$ denotes year-to-year percentage growth in STFP, where α_k, ..., α_s denote Törnqvist input-cost shares and Δk, ..., Δs denote year-to-year percentage growth in observed KLEMS inputs. We compare OTFP and STFP as $\%\Delta OTFP$ and $\%\Delta STFP$ because percentage growth rates abstract from trends and better reveal differences in the two TFPs. Let $r_t = (r_{1t}, \ldots, r_{4t})^T = (\Delta p_{2t} - \Delta p_{1t}, \ldots, \Delta p_{5t} - \Delta p_{1t})^T$ denote a 4x1 vector of differences in percentage growth rates of KLEMS input prices and let $||r_t|| = \sqrt{\sum_{i=1}^{4} r_{it}^2}$ denote the Euclidian norm of r_t. When relative input prices change significantly, $||r_t||$, the 2nd- to 4th-order terms in approximate optimal input function (4.1) which underlies OTFP, and $\%\Delta OTPF - \%\Delta STFP$ are all significantly nonzero. Figure 5a graphs $\%\Delta OTPF - \%\Delta STFP$ and figure 5b graphs $||r_t||$, from 1949 to 2001.

Figure 5a shows a slightly negative average, a slightly upward trend, a relatively large variance from 1949 to the mid 1970s, a declining variance over the whole period, and a relatively small variance from the mid 1970s to 2001. Table 2 summarizes the distribution of $\%\Delta OTFP - \%\Delta STFP$ in figure 5a: minimum = -.013, maximum = .019, average a = -.001, and standard deviation s = .006. A negative average is expected because $\%\Delta STFP$ is based on nonoptimal inputs. Confidence intervals increase in absolute value when translated to levels. For example, if OTPF and STFP are both normalized to one in 1949 and over the 53 years $a-s = -.007 \leq \%\Delta OTFP - \%\Delta STFP \leq a+s = .005$, then, based on a normal distribution, with 68% probability, in 2001, $|OTFP - STFP| \leq .01$

$\%\Delta OTFP - \%\Delta STFP$ has frequently been significant relative to the average values of $\%\Delta OTFP$ and $\%\Delta STFP$. From 1949 to 2001, average $\%\Delta OTFP$ = .0112, average $\%\Delta STFP$ = .0114, and $|\%\Delta OTFP - \%\Delta STFP| > .01$ in 8 out of the first 26 years in the period. Thus, relative to the average values of $\%\Delta OTFP$ and $\%\Delta STFP$, $|\%\Delta OTFP - \%\Delta STFP|$ exceeded about 100% about 30% of the time in the first half of the period. According to equation (4.4), large values of $|\%\Delta OTFP - \%\Delta STFP|$ are caused by large values of $||r_t||$. This appears to be the case somewhat from 1949 to the mid 1970s, strongly in the mid 1970s, but not so much thereafter. Thus,

although the average value of %ΔOTFP - %ΔSTFP has been small from 1949 to 2001 (about .001), in certain years in the first part of the period, |%ΔOTFP - %ΔSTFP| has been large. We conclude that STFP has frequently been significantly biased relative to %ΔOTFP, but not systematically.

An economic argument questions whether any residual can correctly measure TFP. The argument is that TFP represents knowledge and technology that accumulate slowly over time, mostly as a result of conscious investment decisions. That is, TFP moves trendlike with very little noisy variation, unlike the residual measures whose noisy variations presumably mostly reflect shorter-term cyclical variations in observed output. This viewpoint, which suggests developing a structural model in which TFP accumulates as a result of endogenous investment decisions, has been implemented by Chen and Zadrozny (2004).

6. Conclusion.

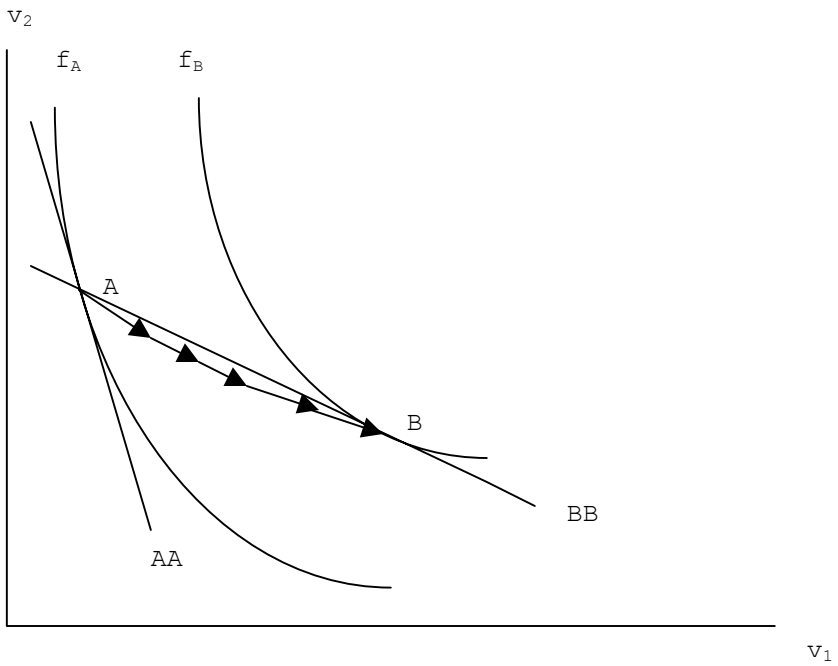
In the paper, we used the multi-step perturbation (MSP) method to estimate CES and TCES production-function models of KLEMS inputs in U.S. manufacturing from 1949 to 2001. For each estimated model, we computed AIC, BCAIC, and BIC and chose as the best one $TCES_1$ model 5 with positive degrees of freedom (DF), which minimized one or more of these information criteria for the sample. By marginally choosing model 5 as the best model, the principal of minimum IC slightly rejects a CES production function, with a single input-price elasticity of substitution for all KLEMS inputs, in favor of a $TCES_1$ model with unitary outer and less than unitary inner input-price elasticities of substitution. Then, we computed the year-to-year percentage growth of optimal TFP (%ΔOTFP) based on the best model and compared it with the percentage growth of standard Solow-residual TFP (%ΔSTFP). Because the model underlying the %ΔSTFP has zero DF, strictly, in contrast to %ΔOTFP, %ΔSTFP should be considered as having no statistical reliability. However, to the extent that %ΔSTFP differs from statistically-reliable %ΔOTFP by less than the average value of %ΔSTFP - %ΔOTFP (about .001), %ΔSTFP can be considered statistically reliable. Nevertheless, relative to the average values of %ΔSTFP and %ΔOTFP, |%ΔSTFP - %ΔOTFP| exceeded about 100% about 30% of the time from 1949 to the mid 1970s. According to Taylor-series theory %ΔOTFP - %ΔSTFP should be accounted for by the 2nd- to 4th-order terms absent from the 1st-order approximate input function (4.3) which underlies %ΔSTFP.

For given estimated input-cost-share parameters, α_i, we estimated elasticity parameters, σ_i, by minimizing -(2/T) × log-likelihood function over a coarse grid of values. In the future, we shall consider estimating the σ_i's more fully over continuous intervals, but still conditional on prior estimates of α_i's because, unless the production function includes a measure of capacity, estimating the α_i's and σ_i's jointly tends to result in implausible estimates of the α_i's (Tatom, 1980). We shall also consider using the more general GCES production function, which in log form is

$$(6.1) \qquad f(v) = (1/\gamma)\cdot\ln\left(\sum_{i=1}^{n} \alpha_i e^{\rho_i v_i}\right),$$

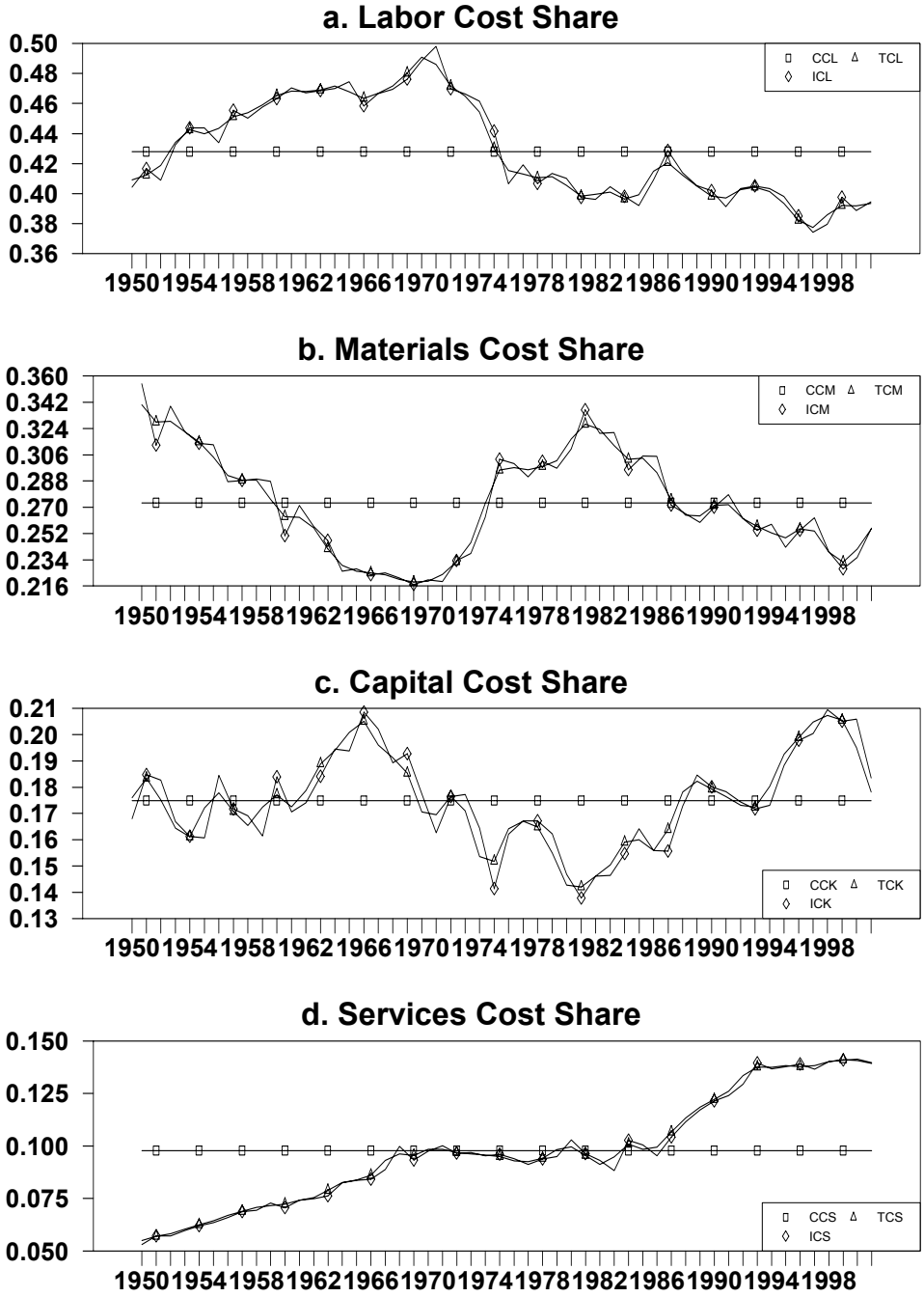
where each input i has its own share parameter ($0 < \alpha_i < 1$) and its own elasticity parameter ($\rho_i < 1$). When the ρ_i's are unequal, the GCES production function is globally nonhomothetic and the 1st-order conditions (2.1) and (2.2) generally have no analytical solution. Because the MSP method produced accurate solutions for the CES and TCES applications here, it should similarly produce accurate solutions for GCES applications.

Figure 1: Illustration of Multi-Step Perturbation.



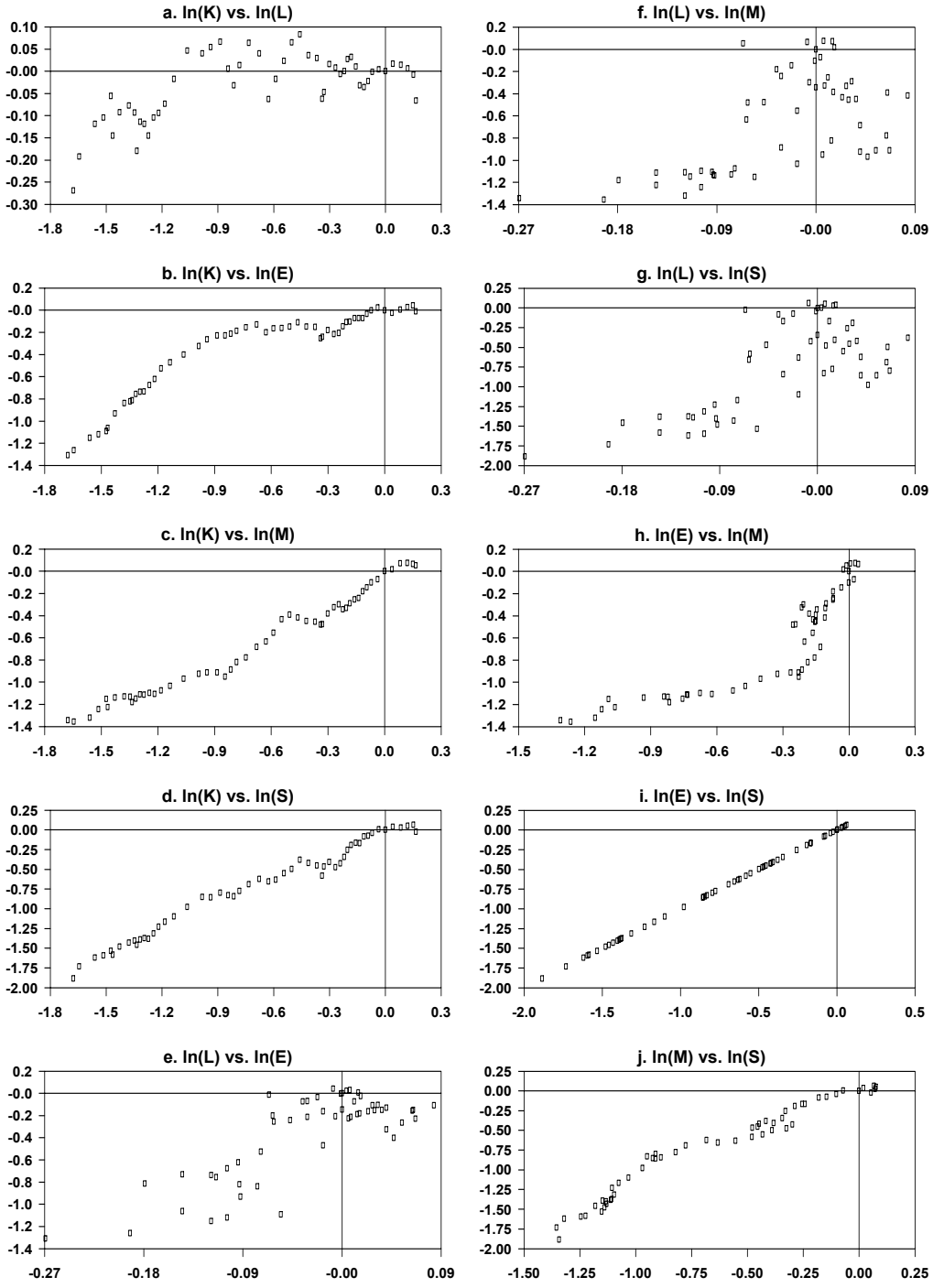
Input-cost lines AA and BB are, respectively, defined by $e(\hat{p})^{T}e(\hat{v}) = e(\hat{p})^{T}e(v)$ and $e(p)^{T}e(\hat{v}) = e(p)^{T}e(v)$, for given precomputed "optimal" $\hat{p}$ and given observed p and v.

Figure 2: Constant, IMA, and Törnqvist LMKS Input Cost Shares, 1949-2001.



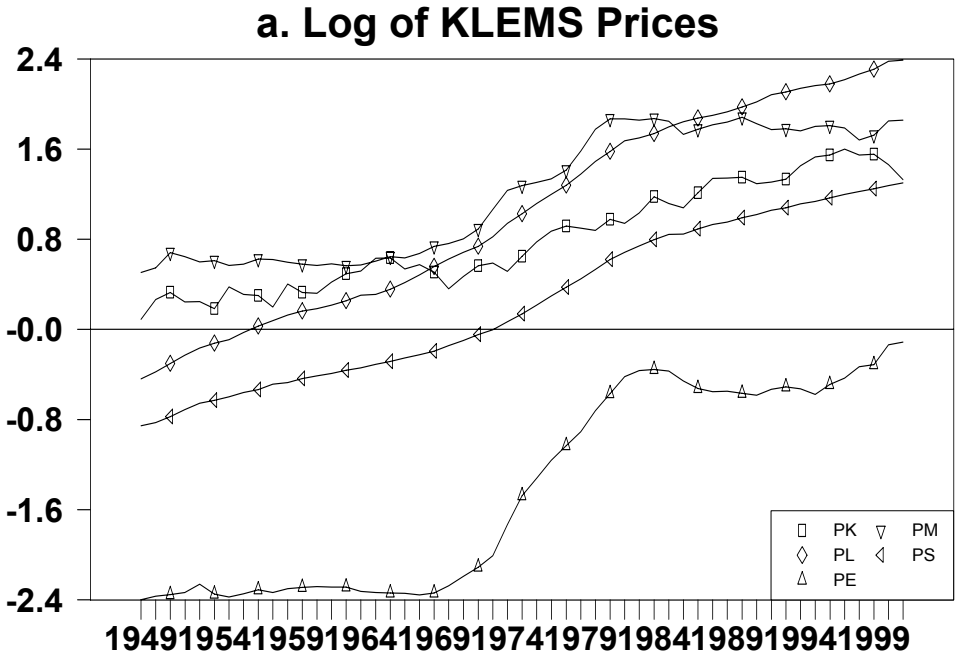
CCL, ... CCS denote constant cost shares of labor, ..., constant cost shares of services; ICL, ..., ICS denote IMA cost shares of labor, ..., IMA cost shares of services; and, TCL, ..., TCS denote Törnqvist cost shares of labor, ..., Törnqvist cost shares of services.

Figure 3: Scatter Plots of Pairwise Log of KLEMS Input Quantities.



The first and second variables listed at the top of each graph refer, respectively, to the vertical and horzontal axes.

Figure 4: Log of KLEMS Input Prices and Quantities, 1949-2001.



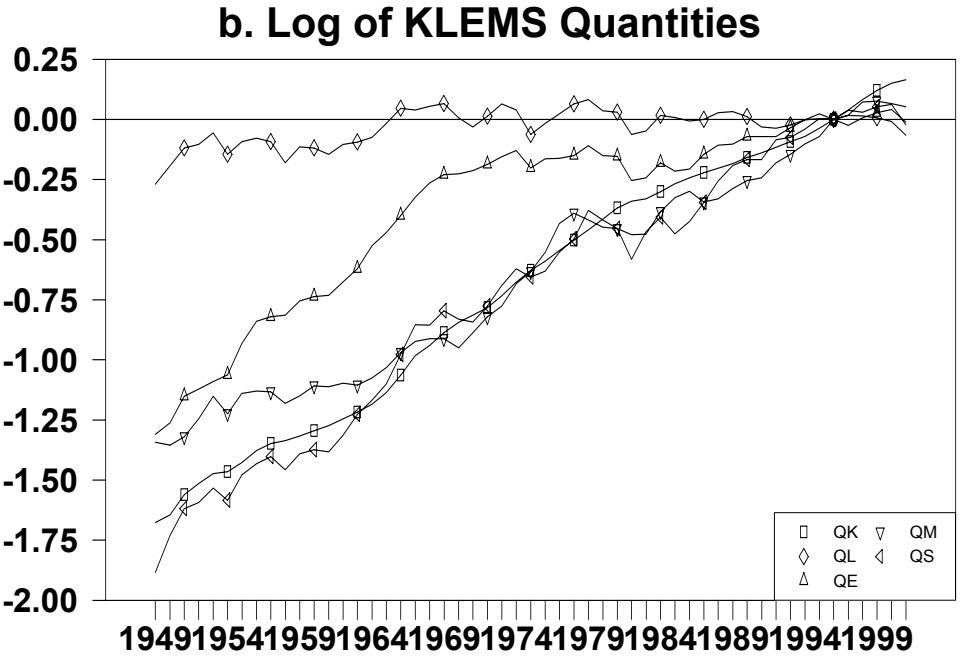
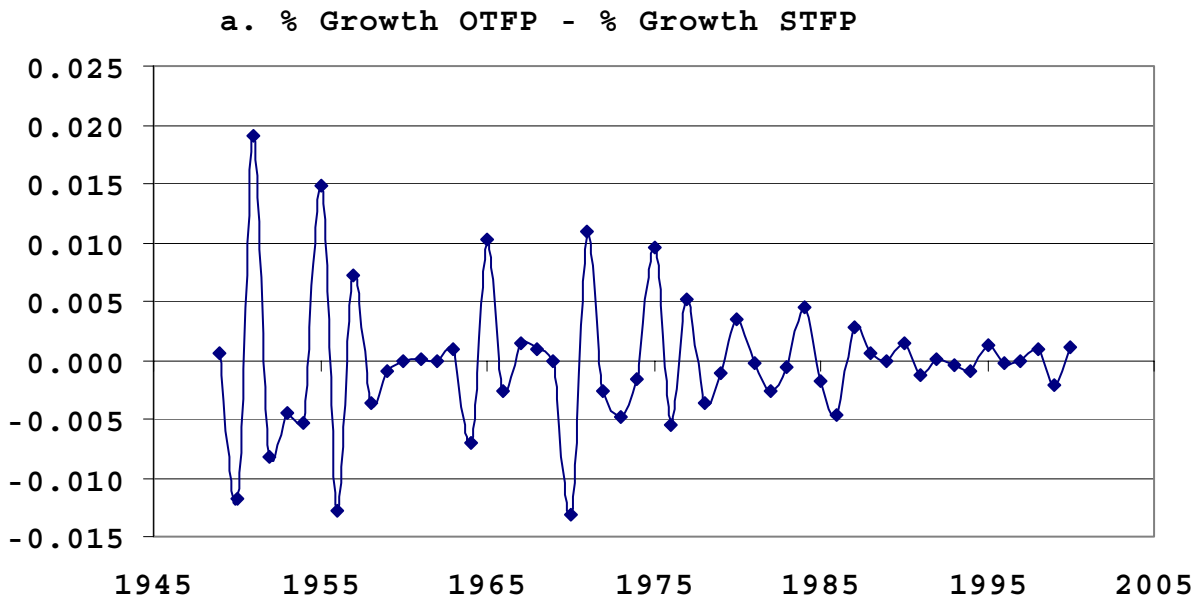


Figure 5: %ΔOTFP - %ΔSTFP and Norms of Differences in %Δ of Relative Input Prices, 1949 to 2001.



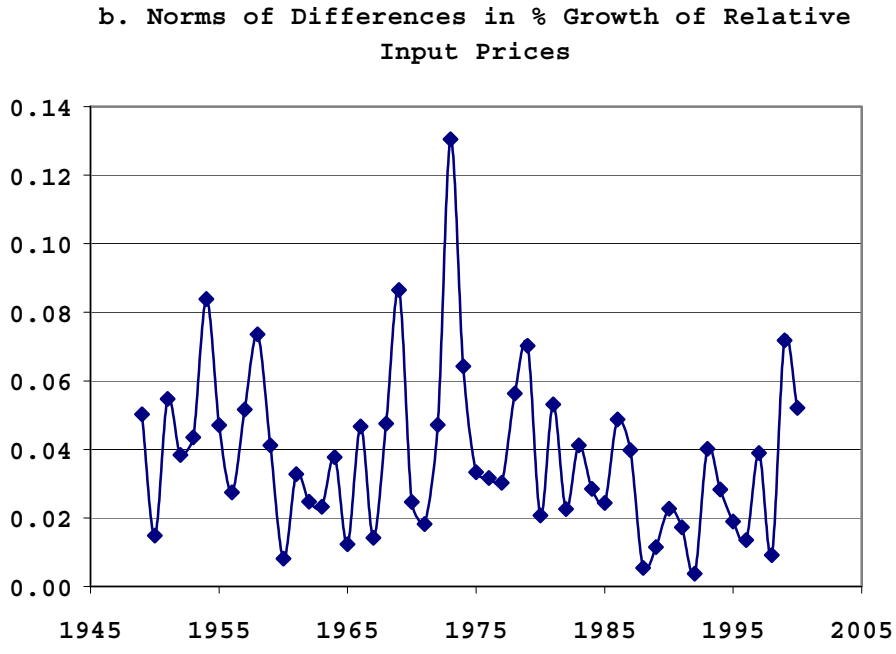


Table 1: Summary Statistics of the Best Estimated Models.

1	2	3	4	5	6	7	8	9
Model	α_{it}	σ_1	σ_2	**DF**	**-(2/T)L**	**AIC**	**BCAIC**	**BIC**
Best CES Models								
1	**Const**	.50	---	47	-26.66	-26.47	-26.45	-26.29
2	**IMA**	1.0	---	43	-32.62	-32.28	-32.21	-31.94
3	**Tornq**	1.0	---	0	-42.22	-34.18	$+\infty$	-26.26
Best $TCES_1$ Models								
4	**Const**	.50	.17	46	-27.69	-27.50	-27.48	-27.32
5	**IMA**	1.0	.67	42	-33.54	-33.21	-33.13	-32.87
6	**Tornq**	1.0	.67	0	-37.26	-26.98	$+\infty$	-19.06
Best $TCES_2$ Models								
7	**Const**	.50	.10	46	-24.09	-23.90	-23.88	-23.71
8	**IMA**	1.0	.67	42	-33.24	-32.90	-32.82	-32.56
9	**Tornq**	1.0	.67	0	-40.35	-32.32	$+\infty$	-24.40

CES, $TCES_1$, and $TCES_2$ production functions are, respectively, $Q = (\alpha_1 K^{\rho} + \alpha_2 L^{\rho} + \alpha_3 E^{\rho} + \alpha_4 M^{\rho} + \alpha_5 S^{\rho})^{1/\rho}$, $Q = [\alpha_1 L^{\rho} + \alpha_2 (\beta_1 K^{\gamma} + \beta_2 E^{\gamma} + \beta_3 M^{\gamma} + \beta_4 S^{\gamma})^{\rho/\gamma}]^{1/\rho}$, and $Q = [\alpha_1 L^{\rho} + \alpha_2 E^{\rho} + \alpha_3 (\beta_1 K^{\gamma} + \beta_2 M^{\gamma} + \beta_3 S^{\gamma})^{\rho/\gamma}]^{1/\rho}$.

Table 2: Summary Statistics of the Distribution of %ΔOTFP - %ΔSTFP.

Min.	-.013
Max.	.019
Mean	-.001
Std. dev.	.006

REFERENCES

Akaike, H. (1973), "Information Theory and Extension of the Maximum Likelihood Principle," pp. 267-281 in Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), Budapest: Akademia Kiado.

Apostol, T.M. (1974), Mathematical Analysis, second edition, Reading, MA: Addison-Wesley.

Arrow, K.J, H.B. Chenery, B.S. Minhas, and R.M. Solow (1961), "Capital-Labor Substitution and Economic Efficiency," Review of Economics and Statistics 43: 225-250.

Berndt, E.R. (1991), The Practice of Econometrics: Classic and Contemporary, Reading, MA: Addison-Wesley.

Bureau of Labor Statistics (2002), Multifactor Productivity web page, http://www.bls.gov/mfp/home.htm.

Burnside, C., M. Eichenbaum, and S. Rebelo (1995), "Capital Utilization and Returns to Scale," Working Paper No. 5125, National Bureau of Economic Research, Cambridge, MA.

Chen, B. and P.A. Zadrozny (2003), "Higher Moments in Perturbation Solution of the Linear-Quadratic Exponential Gaussian Optimal Control Problem," Computational Economics 21: 45-64.

Chen, B. and P.A. Zadrozny (2004), "Endogenous Technical Change in U.S. Manufacturing from 1947 to 1997," working paper, Bureau of Labor Statistics, Washington, DC.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1971), "Conjugate Duality and the Transcendental Logarithmic Production Function," Econometrica 39: 255-256.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1973), "Transcendental Logarithmic Production Functions," Review of Economics and Statistics 55: 28-45.

Gardner, E.S. (1985), "Exponential Smoothing: The State of the Art," Journal of Forecasting 4: 1-28.

Golub, G.H. and C.F. Van Loan (1996), Matrix Computations, third edition, Baltimore, MD: Johns Hopkins University Press.

Hurvich, C.M. and C.L. Tsai (1989), "Regression and Time Series Model Selection in Small Samples," Biometrika 76: 297-307.

Mann, H.B. (1943), "Quadratic Forms with Linear Constraints," American Mathematical Monthly 7: 430-433.

Samuelson, P.A. (1947), Foundations of Economic Analysis, Cambridge, MA: Harvard University Press.

Sato, K. (1967), "A Two-Level Constant Elasticity-of-Substitution Production Function," Review of Economic Studies 34: 201-218.

Schwarz, G. (1978), "Estimating the Dime nsion of a Model," Annals of Statistics 8: 147-164.

Solow, R.M. (1957), "Technical Change and the Aggregate Production Function," Review of Economics and Statistics 39: 312-320.

Tatom, J.A. (1980), "The 'Problem' of Procyclical Productivity," Journal of Political Economy 88: 385-394.

Vartia, Y. (1983), "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions," Econometrica 51: 79-98.

Zadrozny, P.A. and B. Chen (2005), "Multi-Step Perturbation Method for Accurately Computing and Empirically Evaluating the Cost of Living," typescript, Bureau of Labor Statistics, Washington, DC.