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.peter@bls.gov

September 30, 2004

Additional key words: index numbers, tiered CES production functions

JEL codes: C32, C43, C53, C63

ABSTRACT

We describe and illustrate a method for computing and evaluating total factor productivity (TFP). First, we describe using the multi-step perturbation (MSP) method to compute TFP based on any k+1-times differentiable production function. We illustrate the method for Cobb-Douglas (CD), constant elasticity of substitution (CES), and tiered CES (TCES) production functions. Thus, we describe and illustrate computing TFP for far more general production functions than the CD production function, which is the basis of the usual Solow-residual computation of TFP. Second, simultaneously, we describe and illustrate a method for choosing the empirically most valid production function and implied TFP. The MSP method computes optimal inputs, hence, input residuals, the differences between observed and optimal inputs. Optimal inputs maximize output, for a given production function, input prices, and input costs. An information criterion (IC) has a log-likelihood term, computed as the determinant of the sample covariance matrix of input residuals, and a parameter-penalty term, which increases with the number of estimated parameters. Usually, the production function -- the model in this case -- which implies the lowest IC is considered the best or empirically-most-valid production function among those being considered. In effect, the usual Solow residual sets all input residuals identically to zero, which results in estimated parameters and implied TFP having no degrees of freedom and no statistical reliability. By contrast, we use the MSP method to compute TFP based on the "best" CD, CES, or TCES model with the lowest IC and compare it with the usual Solow-residual TFP. We do this using a sample of data on capital, labor, energy, materials, and services (KLEMS) inputs, from 1949-2001, obtained from the Bureau of Labor Statistics.

---

*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.

1. Introduction.

We describe and illustrate a method for computing and evaluating total factor productivity (TFP). First, we describe using the multi-step perturbation (MSP) method to compute TFP based on any k+1-times differentiable production function. Second, simultaneously, we describe and illustrate a method for choosing the empirically most valid production function and implied TFP. We could equally well proceed in terms of variables in original or natural logarithmic (log) form, but proceed in log form for four reasons: (i) total factor productivity (TFP) and related price and quantity indexes are usually considered in log form as period-to-period percentage changes; (ii) log-form variables are unit free, scaled equivalently, hence, lie mostly within or close to a unit sphere, which promotes numerical accuracy; (iii) derivatives of log-form production functions are simpler to derive, program, and compute with; and, (iv) comparisons with the benchmark Solow residuals are easier in log form.

Let q denote the log of the quantity of observed goods and services, let $f(\cdot)$ denote the log of output produced by an assumed k-times differentiable production function, and, let τ denote the log of the level of technology or TFP of $f(\cdot)$. Following Solow (1957), let $\tau = q - \hat{q}$, where $\hat{q} = f(\hat{v})$ denotes the log of optimal output produced by optimal inputs, $\hat{v}$. 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. 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 k+1-order Taylor series, although in the paper we set k = 4. 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 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 τ is absent from the statement of the problem, it plays no role in its solution. Like Solow, we compute τ residually: first $\hat{v}$, then τ.

We consider only interior solutions which satisfy the usual first-order conditions (2.1) and (2.2). As functional forms, we consider Cobb-Douglas (CD) functions, constant elasticity of substitution (CES) functions, and more general tiered CES (TCES) functions, which are multi-level generalizations of two-level CES functions (Sato, 1967; Burnside, Eichenbaum, and Rebelo, 1995), that allow different input groupings to have different substitution elasticities. For each production function, we solve for optimal inputs using the multi-step perturbation (MSP) method, which was developed for the mathematically equivalent utility maximization problem (Zadrozny, 2004). Whereas CD and CES cases always have analytical solutions, generally, TCES cases have only numerical solutions computed using the MSP method or some other numerical method. We use analytical solutions in the CD and CES cases to check the MSP method's accuracy. Independently of the present application, the MSP method is a fairly general numerical method for accurately and quickly solving nonlinear algebraic equations, which are differentiable multiple times.

By a model we mean (i) a k+1-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 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 period-by-period 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 fixed over a sample in terms of the structural parameters and are statistically reliable 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 consider the in-

between case (b), in which nT reduced-form parameters vary smoothly according to an integrated moving-average process (Gardner, 1985), such that $\dim(\theta) < nT$.

Whereas Solow-residual TFP is based on a first-order CD approximation of a differentiable production function, we also consider the more general CES and TCES production functions. However, what difference does the extra generality make? Normally, empirical validity is measured by residual size. In this case, we have output residuals, $q - \hat{q}$, and input residuals, $v - \hat{v}$, where v denotes observed inputs. However, because TPF 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. An IC equals -2 times the log-likelihood function of the residuals plus a penalty term for the uncertainty of estimating parameters. The many IC which have been proposed differ in their propensities for choosing models with particular numbers of parameters. For example, Akaike's IC (1973) tends to pick less-parsimonious models (i.e., with more parameters), Schwarz's IC (1978) tends to pick more parsimonious models, and Hannan and Quinn's IC (1979) tends to pick models with intermediate parsimony. Manzan's (2002) applications suggest that Hannan and Quinn's IC could generally be a safer in-between criterion.

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-share parameters period-by-period to relative input costs, 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 using the MSP method and testing with an IC based on input residuals, we can select the empirically-most-valid model among CD, 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-numerical solutions in the CD and CES cases. We illustrate these ideas using annual data on capital, labor, energy, materials, and services (KLEMS) inputs in manufacturing industries, from 1949 to 2001, obtained from the Bureau of Labor Statistics. Thus, we provide a new method for computing the "empirically-most-valid" TFP, potentially more valid than the Solow residual. Equivalently, we check the robustness of Solow residuals to deviations of the production function from the CD approximation which underlies the Solow residuals.

The paper is organized as follows. Section 2 describes the MSP method for computing optimal and residual inputs. Section 3 does three things. First, it applies the MSP method to the KLEMS data to compute input residuals for the CD, CES, and TCES models being considered. Second, it applies the Hannan and Quinn IC to select a best model among those being considered. Third, for the KLEMS data, it computes TFP implied by the best model and by the benchmark Solow residual and compares the two TFP measures. Section 4 contains concluding remarks.

2. MSP Computation of Total Factor Productivity.

2.1. Preliminary Discussion.

In section 3, we exploit an important property for simplifying the computations, namely that maximizing a function in a constraint set results in a solution which is equivalent to the solution obtained by maximizing a monotonic transformation of the function in the same constraint set. In original units of measurement, 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 first-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 CD, 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, $v_t - \hat{v}_t$. We now describe MSP computation of $\hat{v}_t$ in terms of the movement in figure 1 from points A to B.

Figure 1 considers two inputs. 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 first-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.

2.2. MSP Computation of Optimal Inputs.

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 first-order conditions of the maximization problem by differentiating ℓ with respect to $\hat{v}$ and $\hat{\lambda}$ and setting the results to zero,

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

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

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 second-order conditions 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 first-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 original units of observed inputs are positive, 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; $e(\hat{p})^T e(v) = e(p)^T e(v)$ 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

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

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

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 first-order conditions (2.1) and (2.2) with respect to $\hat{v}$, $\hat{\lambda}$, and p and write the result as

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

or

$$\begin{bmatrix} \nabla^2 f(\hat{v}) - \hat{\lambda} E(p + \hat{v}) & -e(p + \hat{v}) \\ -e(p + \hat{v})^T & 0_{1x1} \end{bmatrix} \begin{bmatrix} d\hat{v} \\ d\hat{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{\lambda} E(p + \hat{v}) \\ e(p + \hat{v})^T - e(p + v)^T \end{bmatrix} dp,$$

where $\nabla^2 f(\cdot)$ denotes the n×n Hessian matrix of second-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 all of x is computed recursively and hats emphasize that values are computed, for simplicity, we omit them from x because, unlike in v or y, we do not need to distinguish between hatted and unhatted x. If $f(\cdot)$ is differentiable k+1 times, then, $F(x)$ is differentiable k-1 times; $G(x)$ is always differentiable any number of times.

The elements of x are all known, because they are either observed or 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 $|F(x)| \neq 0$, where $|\cdot|$ denotes the determinant of a square matrix. This condition holds because the second-order conditions of the problem imply that

$$(2.5) \qquad (-1)^{n+1}|F(x)| > 0$$

(Mann, 1943). Thus, when x maximizes output and satisfies second-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, ..., 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, upon which the MSP method is based, implies that if the production function is differentiable k+2 times and satisfies the second-order conditions, so that its optima are interior points, then, the exact solution path is differentiable k+1 times, and, in each computational subperiod $s \in [t_i, t_{i+1})$, for i = 1, ..., h, has the kth-order polynomial Taylor-series approximation of the true y(s),

$$\hat{y}(s) = \hat{y}_{t_i} + \nabla\hat{y}_{t_i}(s-t_i) + (1/2)\nabla^2\hat{y}_{t_i}(s-t_i)^2 + \ldots + (1/k!)\nabla^k\hat{y}_{t_i}(s-t_i)^k, \tag{2.7}$$

where $\nabla\hat{y}_{t_i}$, ..., $\nabla^k\hat{y}_{t_i}$ are (n+1)×1 coefficients to be computed in terms of observed input prices and quantities. We state an analogous polynomial price process in equation (2.15). See Ford (1955), for example, for a discussion which similarly connects continuous- and discrete-time processes through polynomial interpolation.

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 $i = 1, \ldots, h$. For each subperiod i in period t, we compute the coefficients of the approximate process (2.7), $\nabla\hat{y}_{t_i}$, ..., $\nabla^k\hat{y}_{t_i}$, and, then, compute the subperiod increments, $\Delta\hat{y}_{t_i}$, as

$$\Delta\hat{y}_{t_i} = \nabla\hat{y}_{t_i} h^{-1} + (1/2)\nabla^2\hat{y}_{t_i} h^{-2} + \ldots + (1/k!)\nabla^k\hat{y}_{t_i} h^{-k}. \tag{2.8}$$

The theoretical approximation error is on the order of h^{-k}, which is denoted by $\varepsilon = O(h^{-k})$. The error can presumably be controlled by setting h and k sufficiently large, although h and k cannot be so large that h^{-k} cannot be stored as a nonzero number on the computer being used. See Zadrozny (2004, table 1) for values of h and k which predict achieving particular orders of magnitude of accuracy. In practice, because numerical computations have finite precision, the theoretical error is augmented by numerical errors. In particular, significant numerical errors can arise when inverting $F(x)$, but these can be controlled by making indifference curves sufficiently curved, namely, by ensuring that $|F(x)|$ is sufficiently far from zero.

We now outline the MSP method for $k = 4$. All differentiations are with respect to continuous time, s. We differentiate the approximate y process (2.7) four times and state the resulting first to fourth differentials in equations (2.9). We differentiate the first differential (2.6) of the true y process three times and state the resulting second to fourth differentials in equations (2.10), in terms of differentials of $H(x)$ and prices. We state the differentials of $H(x)$ in terms of differentials of $F(x)$ and $G(x)$, in equations (2.11), and state the latter in equations (2.12) to (2.14). We assume the 4th-order polynomial price process (2.15), state its first to fourth differentials in equations (2.16), and state its coefficients, ∇p_t, ..., $\nabla^4 p_t$, in terms of backward differences of prices, in equations (2.17) and (2.18). Then, for each period t and subperiod i, we combine the results, compute the coefficients, $\nabla\hat{y}_{t_i}$, ..., $\nabla^k\hat{y}_{t_i}$, of the approximate process (2.7), according to equations (2.19) to (2.34), and compute the subperiod increments, $\Delta\hat{y}_{t_i}$, according to equation (2.8). Finally, we obtain the period increment as the sum $\Delta\hat{y}_t = \sum_{i=1}^{h} \Delta\hat{y}_{t_i}$.

For $k = 4$, $i = 1, \ldots, h$, and $s \in [t_i, t_i+h)$, we differentiate the approximate y process (2.7) with respect to s and obtain

$$(2.9) \qquad d\hat{y}(s) = \nabla\hat{y}_{t_i} + \nabla^2\hat{y}_{t_i}(s-t_i) + (1/2)\nabla^3\hat{y}_{t_i}(s-t_i)^2 + (1/6)\nabla^4\hat{y}_{t_i}(s-t_i)^3.$$

$$d^2\hat{y}(s) = \nabla^2\hat{y}_{t_i} + \nabla^3\hat{y}_{t_i}(s-t_i) + (1/2)\nabla^4\hat{y}_{t_i}(s-t_i)^2,$$

$$d^3\hat{y}(s) = \nabla^3\hat{y}_{t_i} + \nabla^4\hat{y}_{t_i}(s-t_i),$$

$$d^4\hat{y}(s) = \nabla^4\hat{y}_{t_i}.$$

We compute the coefficients, $\nabla\hat{y}_{t_i}$, ..., $\nabla^4\hat{y}_{t_i}$, so that they are equal to the first to fourth differentials of the true y process (2.6).

Using the product rule of differentiation, we differentiate equation (2.6) three times and obtain

$$(2.10) \qquad d^2\hat{y}(s) = dH(s)dp(s) + H(s)d^2p(s),$$

$$d^3\hat{y}(s) = d^2H(s)dp(s) + 2dH(s)d^2p(s) + H(s)d^3p(s),$$

$$d^4\hat{y}(s) = d^3H(s)dp(s) + 3d^2H(s)d^2p(s) + 3dH(s)d^3p(s) + H(s)d^4p(s),$$

where $H(s) \equiv F(x(s))^{-1}G(x(s))$.

Repeatedly applying the product rule of differentiation to $F(s)H(s) = G(s)$, we obtain $dF(s)H(s) + F(s)dH(s) = dG(s)$, $d^2F(s)H(s) + 2dF(s)dH(s) + F(s)d^2H(s) = d^2G(s)$, and $d^3F(s)H(s) + 3d^2F(s)dH(s) + 3dF(s)d^2H(s) + F(s)d^3H(s) = d^3G(s)$, where $d^kF(s) \equiv d^kF(x(s))$ and $d^kG(s) = d^kG(x(s))$, for $k = 0, \ldots, 3$. Then, solving for $dH(s)$, $d^2H(s)$, and $d^3H(s)$, we obtain

$$(2.11) \qquad dH(s) = F(s)^{-1}[dG(s) - dF(s)H(s)],$$

$$d^2H(s) = F(s)^{-1}[d^2G(s) - d^2F(s)H(s) - 2dF(s)dH(s)],$$

$$d^3H(s) = F(s)^{-1}[d^3G(s) - d^3F(s)H(s) - 3d^2F(s)dH(s) - 3dF(s)d^2H(s)].$$

Equations (2.11) are fully recursive. Given the production function, $f(\cdot)$, we compute $F(s)$, $G(s)$, and $H(s)$. Then, we compute $dF(s)$, $dG(s)$, and $dH(s)$. Then,

we compute $d^2F(s)$, $d^2G(s)$, and $d^2H(s)$. Finally, we compute $d^3F(s)$, $d^3G(s)$, and $d^3H(s)$. For k = 1, 2, and 3, $d^kF(s)$ and $d^kG(s)$ have the form

$$(2.12) \qquad d^kF(s) = \begin{bmatrix} d^k(\nabla^2 f(\hat{v})) - d^k(\hat{\lambda}E(p + \hat{v})) & - d^k e(p + \hat{v}) \\ - d^k e(p + \hat{v})^T & 0_{1x1} \end{bmatrix},$$

$$d^kG(s) = \begin{bmatrix} d^k(\hat{\lambda}E(p + \hat{v})) \\ d^k e(p + \hat{v})^T - d^k e(p + v)^T \end{bmatrix}.$$

where p and $\hat{v}$ are functions of s and v is constant.

Using the definitions of matrix derivatives and rules of matrix differentiation in Chen and Zadrozny (2003), we differentiate $\nabla^2 f(\hat{v}(s))$ three times with respect to s and obtain

$$(2.13) \qquad \mathrm{vec}(d(\nabla^2 f(\hat{v}(s)))) = \nabla^3 f(\hat{v}(s)) d\hat{v}(s),$$

$$\mathrm{vec}(d^2(\nabla^2 f(\hat{v}(s)))) = [d\hat{v}(s)^T \otimes I_n]\nabla^4 f(\hat{v}(s)) d\hat{v}(s) + \nabla^3 f(\hat{v}(s)) d^2\hat{v}(s),$$

$$\mathrm{vec}(d^3(\nabla^2 f(\hat{v}(s)))) = [d\hat{v}(s)^T \otimes d\hat{v}(s)^T \otimes I_{n^2}]\nabla^5 f(\hat{v}(s)) d\hat{v}(s)$$

$$+ 2[d^2\hat{v}(s)^T \otimes I_{n^2}]\nabla^4 f(\hat{v}(s)) d\hat{v}(s)$$

$$+ \nabla^3 f(\hat{v}(s)) d^3\hat{v}(s),$$

where I_n and I_{n^2} are nxn and n^2xn^2 identity matrices and $d^k\hat{v}(s)$ is the top n-dimensional subvector of $d^k\hat{y}(s)$. By definition, $\nabla f(\hat{v}) = 1\times n = [\partial f(\hat{v})/\partial v_1, \ldots, \partial f(\hat{v})/\partial v_n]$, $\nabla^2 f(\hat{v}) = n\times n = [\partial \mathrm{vec}(\nabla f(\hat{v}))/\partial v_1, \ldots, \partial \mathrm{vec}(\nabla f(\hat{v}))/\partial v_n]$, $\ldots$, $\nabla^5 f(\hat{v}) = n^4\times n = [\partial \mathrm{vec}(\nabla^4 f(\hat{v}))/\partial v_1, \ldots, \partial \mathrm{vec}(\nabla^4 f(\hat{v}))/\partial v_n]$. In section 3, for the CD, CES, and TCES production functions, we actually use simpler equivalent equations, without Kronecker products.

To compute $d^kF(s)$ and $d^kG(s)$, we use the product rule of differentiation and obtain $d(\hat{\lambda}E(p+\hat{v})) = d\hat{\lambda}\cdot E(p+\hat{v}) + \hat{\lambda}\, dE(p+\hat{v})$, $d^2(\hat{\lambda}E(p+\hat{v})) = d^2\hat{\lambda}\cdot E(p+\hat{v}) +$

$2d\hat{\lambda}\cdot dE(p+\hat{v}) + \hat{\lambda}d^2E(p+\hat{v})$, and $d^3(\hat{\lambda}E(p+\hat{v})) = d^3\hat{\lambda}\cdot E(p+\hat{v}) + 3d^2\hat{\lambda}\cdot dE(p+\hat{v}) + 3d\hat{\lambda}\cdot d^2E(p+\hat{v}) + \hat{\lambda}d^3E(p+\hat{v})$, where

(2.14) $de(p+\hat{v}) = E(p+\hat{v})(dp+d\hat{v})$,

$dE(p+\hat{v}) = E(p+\hat{v})D(dp+d\hat{v})$,

$d^2e(p+\hat{v}) = E(p+\hat{v})[D(dp+d\hat{v})(dp+d\hat{v}) + (d^2p+d^2\hat{v})]$,

$d^2E(p+\hat{v}) = E(p+\hat{v})[D(dp+d\hat{v})^2 + D(d^2p+d^2\hat{v})]$,

$d^3e(p+\hat{v}) = E(p+\hat{v})\{[D(d^2p+d^2\hat{v}) + D(dp+d\hat{v})^2](dp+d\hat{v})$

$+ 2D(dp+d\hat{v})(d^2p+d^2\hat{v}) + (d^3p+d^3\hat{v})\}$,

$d^3E(p+\hat{v}) = E(p+\hat{v})[D(dp+d\hat{v})^3 + 3D(dp+d\hat{v})D(d^2p+d^2\hat{v}) + D(d^3p+d^3\hat{v})]$,

and $D(x)$ denotes the $n\times n$ diagonal matrix with $n\times 1$ vector x on its principal diagonal. In the bottom half of $d^kG(s)$, the v without the hat is constant, so that $d^ke(p+v)$ is evaluated as $d^ke(p+\hat{v})$, according to equations (2.14), with $d^k\hat{v} = 0$, for $k = 1$, 2, and 3.

Analogous to approximate y process (2.7), we assume prices follow a 4th-order polynomial process, for $s \in [t,t+1)$ and $t = 1, \ldots, T$,

(2.15) $p(s) = \hat{p}_t + \nabla p_t(s-t) + (1/2)\nabla^2 p_t(s-t)^2 + (1/6)\nabla^3 p_t(s-t)^3 + (1/24)\nabla^4 p_t(s-t)^4$,

with $n\times 1$ coefficients, $\hat{p}_t$, ∇p_t, ..., $\nabla^4 p_t$. Although the price coefficients remain at their initial values, indexed at $t_1 = t$, throughout computations in period t, the y coefficients, $\nabla\hat{y}_{t_i}$, ..., $\nabla^4\hat{y}_{t_i}$, are indexed by t_i and updated at each iteration i within period t. From price process (2.15), 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. We allow within-period prices to vary nonlinearly on the presumption that in doing so the computations will more accurately capture

the nonlinear effects of input price variations on optimal inputs. Although we could use just a linear price process, with $\nabla^j p_t \equiv 0$ for $j \geq 2$, we expect computations based on a nonlinear price process to be more accurate, because prices affect optimal inputs nonlinearly.

In the following derivations, we distinguish between differentiating with respect to s and differencing with respect to t. For $s \in [t,t+1)$, differentiating price process (2.15) four times with respect to s, we obtain

$$(2.16) \quad dp(s) = \nabla p_t + \nabla^2 p_t(s-t) + (1/2)\nabla^3 p_t(s-t)^2 + (1/6)\nabla^4 p_t(s-t)^3,$$

$$d^2p(s) = \nabla^2 p_t + \nabla^3 p_t(s-t) + (1/2)\nabla^4 p_t(s-t)^2,$$

$$d^3p(s) = \nabla^3 p_t + \nabla^4 p_t(s-t),$$

$$d^4p(s) = \nabla^4 p_t.$$

Then, differencing price process (2.15) backwards in time four times with respect to t, we obtain the price coefficients, ∇p_t, ..., $\nabla^4 p_t$, in terms of backward-differenced prices, Δp_t, ..., $\Delta^4 p_t$, as

$$(2.17) \quad \nabla p_t = \Delta p_t - (1/2)\Delta^2 p_t + (1/3)\Delta^3 p_t - (1/4)\Delta^4 p_t,$$

$$\nabla^2 p_t = \Delta^2 p_t - \Delta^3 p_t + (11/12)\Delta^4 p_t,$$

$$\nabla^3 p_t = \Delta^3 p_t - (3/2)\Delta^4 p_t,$$

$$\nabla^4 p_t = \Delta^4 p_t,$$

where the backward-differenced prices are given in terms of the computed and observed prices, p_t, $\hat{p}_t$, p_{t-1}, p_{t-2}, and p_{t-3}, as

$$(2.18) \quad \Delta p_t = p_t - \hat{p}_t,$$

$$\Delta^2 p_t = p_t - 2\hat{p}_t + p_{t-1},$$

$$\Delta^3 p_t = p_t - 3\hat{p}_t + 3p_{t-1} - p_{t-2},$$

$$\Delta^4 p_t = p_t - 4\hat{p}_t + 6p_{t-1} - 4p_{t-2} + p_{t-3}.$$

The observed prices, p_t, p_{t-1}, p_{t-2}, and p_{t-3}, are automatically in backward-in-time order, while the computed prices, $\hat{p}_t$, are assumed to occur between p_{t-1} and p_t. As required, the price coefficients set by equations (2.17) and (2.18) imply that the price process (2.15) passes through the computed and observed prices. We could achieve the same perfect fit of computed and observed prices using centered or forward differences. We use equations (2.15) to interpolate prices for purely numerical reasons -- not statistical reasons -- so that the perfect fit is coincidental.

We now describe in detail computing $\{\hat{v}_t\}_{t=1}^{T}$ using the MSP method. We can generally guarantee a desired accuracy by setting h and k sufficiently large. The method is quick because the computations are mostly recursive, linear when nonrecursive, and involve a finite total number of computations, namely, do not involve iterations until convergence at any point. 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 seven steps. Within the seven steps, we take as given an assumed $f(\cdot)$ and observed p_t and v_t.

Step 1: Initialize x_t, 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 (2.16), we set the price differentials as $dp(t) = \nabla p_t$, ..., $d^4p(t) = \nabla^4 p_t$. Following equations (2.17) and (2.18), we compute the price coefficients, ∇p_t, ..., $\nabla^4 p_t$, in terms of the computed and observed prices, $\hat{p}_t$, p_t, p_{t-1}, p_{t-2}, and p_{t-3}.

Step 2: Compute 1st-Order y Coefficient.

For $s = t_1 = t$, equation (2.6), the first equation of (2.9), and the first equation of (2.16), imply that

$$(2.19) \qquad H(x_t) = F(x_t)^{-1}G(x_t),$$

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

Step 3: Compute 2nd-Order y Coefficient.

The first equation of (2.13) implies that

$$(2.20) \qquad \mathrm{vec}(d(\nabla^2 f(\hat{v}_t))) = \nabla^3 f(\hat{v}_t)\nabla\hat{v}_t,$$

where $\nabla\hat{v}_t$ is the top n-dimensional subvector of $\nabla\hat{y}_t$. Equation (2.12) implies that

$$(2.21) \qquad dF(x) = \begin{bmatrix} d(\nabla^2 f(\hat{v})) - d(\hat{\lambda}E(p+\hat{v})) & -de(p+\hat{v}) \\ -de(p+\hat{v})^T & 0_{1x1} \end{bmatrix},$$

$$dG(x) = \begin{bmatrix} d(\hat{\lambda}E(p+\hat{v})) \\ de(p+\hat{v})^T - de(p+v)^T \end{bmatrix},$$

where the details of dF(x) and dG(x) are given by equations (2.13) and (2.14). The first equation of (2.11) implies that

$$(2.22) \qquad dH(x_t) = F(x_t)^{-1}[dG(x_t) - dF(x_t)H(x_t)].$$

The second equation of (2.9), the first equation of (2.10), and the first and second equations of (2.16), imply that

$$(2.23) \qquad \nabla^2\hat{y}_t = dH(x_t)\nabla p_t + H(x_t)\nabla^2 p_t.$$

Step 4: Compute 3rd-Order y Coefficient.

The second equation of (2.13) implies that

$$(2.24)\qquad \mathrm{vec}(d^2(\nabla^2 f(\hat{v}_t))) = [\nabla \hat{v}_t^T \otimes I_{n^2}]\nabla^4 f(\hat{v}_t)\nabla\hat{v}_t + \nabla^3 f(\hat{v}_t)\nabla^2\hat{v}_t,$$

where $\nabla^2\hat{v}_t$ is the top n-dimensional subvector of $\nabla^2\hat{y}_t$. Equation (2.12) implies that

$$(2.25)\qquad d^2F(x) = \begin{bmatrix} d^2(\nabla^2 f(\hat{v})) - d^2(\hat{\lambda}E(p+\hat{v})) & -d^2e(p+\hat{v}) \\ -d^2e(p+\hat{v})^T & 0_{1x1} \end{bmatrix},$$

$$d^2G(x) = \begin{bmatrix} d^2(\hat{\lambda}E(p+\hat{v})) \\ d^2e(p+\hat{v})^T - d^2e(p+v)^T \end{bmatrix},$$

where the details of $d^2F(x)$ and $d^2G(x)$ are given by equations (2.13) and (2.14). The second equation of (2.11) implies that

$$(2.26)\qquad d^2H(x_t) = F(x_t)^{-1}[d^2G(x_t) - d^2F(x_t)H(x_t) - 2dF(x_t)dH(x_t)].$$

The third equation of (2.9), the second equation of (2.10), and the first to third equations of (2.16), imply that

$$(2.27)\qquad \nabla^3\hat{y}_t = d^2H(x_t)\nabla p_t + 2dH(x_t)\nabla^2 p_t + H(x_t)\nabla^3 p_t.$$

Step 5: Compute 4th-Order y Coefficient and Update y.

The third equation of (2.13) implies that

$$(2.28)\qquad \mathrm{vec}(d^3(\nabla^2 f(\hat{v}_t))) = [\nabla \hat{v}_t^T \otimes \nabla \hat{v}_t^T \otimes I_{n^2}]\nabla^5 f(\hat{v}_t)\nabla\hat{v}_t$$

$$+ 2[\nabla^2 \hat{v}_t^T \otimes I_{n^2}]\nabla^4 f(\hat{v}_t))\nabla\hat{v}_t + \nabla^3 f(\hat{v}_t)\nabla^3\hat{v}_t,$$

where $\nabla^3\hat{v}_t$ is the top n-dimensional subvector of $\nabla^3\hat{y}_t$. Equation (2.12) implies that

$$(2.29)\qquad d^3F(x) = \begin{bmatrix} d^3(\nabla^2 f(\hat{v})) - d^3(\hat{\lambda}E(p+\hat{v})) & -d^3e(p+\hat{v}) \\ -d^3e(p+\hat{v})^T & 0_{1x1} \end{bmatrix},$$

$$d^3G(x) = \begin{bmatrix} d^3(\hat{\lambda}E(p+\hat{v})) \\ d^3e(p+\hat{v})^T - d^3e(p+v)^T \end{bmatrix},$$

where the details of $d^3F(x)$ and $d^3G(x)$ are given by equations (2.13) and (2.14). The third equation of (2.11) implies that

$$(2.30)\qquad d^3H(x_t) = F(x_t)^{-1}[d^3G(y_t) - d^3F(x_t)H(x_t) - 3d^2F(x_t)dH(x_t) - 3dF(x_t)d^2H(x_t)].$$

The fourth equation of (2.9), the third equation of (2.10), and the first to fourth equations of (2.16), imply that

$$(2.31)\qquad \nabla^4\hat{y}_t = d^3H(x_t)\nabla p_t + 3d^2H(x_t)\nabla^2 p_t + 3dH(x_t)\nabla^3 p_t + H(x_t)\nabla^4 p_t.$$

For k = 4 and i = 1, equation (2.8) implies that

$$(2.32)\qquad \Delta\hat{y}_t = \nabla\hat{y}_t h^{-1} + (1/2)\nabla^2\hat{y}_t h^{-2} + (1/6)\nabla^3\hat{y}_t h^{-3} + (1/24)\nabla^4\hat{y}_t h^{-4},$$

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

Step 6: Update Price Coefficients, x_{t_2}, and y.

For i = 2 and, hence, for $s = t_2 = t+h^{-1}$, equations (2.15) and (2.16) imply that we update prices and their differentials as

$$(2.33)\qquad p(t_2) = \hat{p}_t + \nabla p_t h^{-1} + (1/2)\nabla^2 p_t h^{-2} + (1/6)\nabla^3 p_t h^{-3} + (1/24)\nabla^4 p_t h^{-4},$$

$$dp(t_2) = \nabla p_t + \nabla^2 p_t h^{-1} + (1/2)\nabla^3 p_t h^{-2} + (1/6)\nabla^4 p_t h^{-3},$$

$$d^2p(t_2) = \nabla^2 p_t + \nabla^3 p_t h^{-1} + (1/2)\nabla^4 p_t h^{-2},$$

$$d^3p(t_2) = \nabla^3 p_t + \nabla^4 p_t h^{-1},$$

$$d^4p(t_2) = \nabla^4 p_t,$$

such that the price coefficients, ∇p_t, ..., $\nabla^4 p_t$, remain at their initial $t_1 = t$ computed values. We set $x_{t_2} = (\hat{y}_{t_2}^T, p_{t_2}^T)^T$. We repeat steps 2 to 5 and, thereby, update the y coefficients to $\nabla\hat{y}_{t_2}$, ..., $\nabla^4\hat{y}_{t_2}$. Following equation (2.8), we compute $\Delta\hat{y}_{t_2} = \nabla\hat{y}_{t_2}h^{-1} + (1/2)\nabla^2\hat{y}_{t_2}h^{-2} + (1/6)\nabla^3\hat{y}_{t_2}h^{-3} + (1/24)\nabla^4\hat{y}_{t_2}h^{-4}$ and $\hat{y}_{t_3} = \hat{y}_{t_2} + \Delta\hat{y}_{t_2}$.

Step 7: Repeat Steps 2 to 6.

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

$$p(t_3) = \hat{p}_t + 2\nabla p_t h^{-1} + 2\nabla^2 p_t h^{-2} + (4/3)\nabla^3 p_t h^{-3} + (2/3)\nabla^4 p_t h^{-4}, \tag{2.34}$$

$$dp(t_3) = \nabla p_t + 2\nabla^2 p_t h^{-1} + 2\nabla^3 p_t h^{-2} + (4/3)\nabla^4 p_t h^{-3},$$

$$d^2p(t_3) = \nabla^2 p_t + 2\nabla^3 p_t h^{-1} + 2\nabla^4 p_t h^{-2},$$

$$d^3p(t_3) = \nabla^3 p_t + 2\nabla^4 p_t h^{-1},$$

$$d^4p(t_3) = \nabla^4 p_t.$$

We set $x_{t_3} = (\hat{y}_{t_3}^T, p_{t_3}^T)^T$. We repeat steps 2 to 5 and update the y coefficients to $\nabla\hat{y}_{t_3}$, ..., $\nabla^4\hat{y}_{t_3}$. We compute $\Delta\hat{y}_{t_3} = \nabla\hat{y}_{t_3}h^{-1} + (1/2)\nabla^2\hat{y}_{t_3}h^{-2} + (1/6)\nabla^3\hat{y}_{t_3}h^{-3} + (1/24)\nabla^4\hat{y}_{t_3}h^{-4}$ 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 sub-vector of the computed $\Delta\hat{y}_{t_h}$.

3. Application Using KLEMS Data.

We now use the MSP method to test robustness of Solow residuals as a measure of TFP, by computing TFP for CD, CES, and TCES models. We use 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. As noted before, it makes no difference whether the prices are in current or constant dollars.

The Solow-residual is based on a first-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. The parameters are input-cost shares, denoted α_1, ..., α_5, and input substitution elasticities, denoted σ in the CD and CES models and σ_1 and σ_2, for $\sigma_1 > \sigma_2$, in the TCES models. For the 53 annual periods, we consider "constant" α's estimated as sample means, "IMA" α's equal to one-period ahead forecasts of estimated IMA(1,1) models of the cost shares, and "Törnqvist" α'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 raw cost-share observations. In each case, because the cost shares must sum to one, we set the α's of the four largest LMKS-cost shares as noted and set the remaining E-cost shares residually, as one minus the sum of the other α's. For CES models, we consider σ = .1, .5, 1., 2., 10. Thus, we do a kind of MLE over a coarse grid of σ's, conditional on estimated α's. In TCES models, we consider two σ's over a similar grid, such that the "outer" one is always larger than the "inner" one. We do not consider joint estimation of parameters, such as MLE, because often this results in implausible α's. For example, until he introduces utilization rates (an extension which is beyond the scope of this paper), Tatom (1980) obtains MLE $\alpha_L > 1$ and $\alpha_K < 0$, which contradicts $0 \leq \alpha \leq 1$.

We evaluate estimated models in terms of information criteria (IC). We consider the basic Akaike IC or AIC, the bias corrected AIC or BCAIC (Hurvich and Tsai, 1989), and the Schwarz (1978) Bayesian IC or BIC. We are especially concerned about degrees of freedom (DF) of estimated parameters and, for a chosen IC, consider as "best" the model which minimizes that IC. We are concerned with DF because a model with zero DF implies that the model's estimated parameters and any derived quantities, such as TFP, have infinite variances and, hence, have no statistical reliability. To varying extents, the

ICs considered here account for DF by adding penalty terms to -(2/T) × log-likelihood function. Among the ICs considered here, in tables 1 to 3, BCAIC most effectively accounts for DF, because it is the only IC which approaches +∞ as DF approach zero from above. Thus, we set BCAIC = +∞ when DF are exhausted. An IC is parsimonious if it selects as "best" the models with the fewest parameters. ICs in Tables 1-3 are ordered in increasing parsimony as AIC, BCAIC, and BIC.

3.1. Results from CES Models

We considered 15 CES models. Table 1 reports ICs of the CES production function for five KLEMS input residuals; constant, IMA, and Törnqvist cost-share processes; and five values of the input elasticity of substitution. Of course, when the elasticity of substitution is one, the CES function reduces to the CD function.

The DF in table 1 are obtained as follows. Each model has five KLEMS inputs. Because the cost shares sum to one, there are four free cost shares in each of the 53 sample periods. Each model also has an elasticity parameter. Thus, constant-cost-share models 1-5 have 4 + 1 = 5 estimated parameters, hence, have DF = max[53-5,0] = 48. Each IMA(1,1)-cost-share model has two estimated parameters, a moving-average coefficient and a white-noise disturbance variance. Thus, IMA-cost-share models 6-10 have 4×2 + 1 = 9 estimated parameters, hence, have DF = max[53-9,0] = 44. Finally, Törnqvist-cost-share models 11-15 have 53×4 + 1 = 213 estimated parameters, hence, have DF = max[53-213,0] = 0. 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.

In the constant-cost share case, $\sigma = .5$ yields the best IC values. The shaded boxes highlight IMA-cost-share model 8, which is the best model, with lowest ICs and positive DF. Because the IMA- and Törnqvist-cost share graphs are nearly identical but IMA DF = 44, whereas Törnqvist DF = 0, we consider the IMA models as better, regardless of IC values. Thus, even if we chose to follow AIC and disregard the other ICs, we would consider IMA model 8 better than

Törnqvist model 13, because model 8 has positive DF, even though model 13's AIC is lower.

3.2. Results from TCES Models

We also considered 24 TCES models. Even if we limit the TCES model search to two-tiered models, this still implies more models than we could evaluate in practice, because there are 16 possible groupings with one to five KLEMS inputs. Thus, we look 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 high substitution σ_1, and an inner group of K, E, M, and S, with low substitution σ_2. The L-KEMS two-tiered CES model, $TCES_1$, takes the form

$$(3.1) \qquad Q = [\alpha_1 L^{\rho} + \alpha_2(\beta_1 K^{\gamma} + \beta_2 E^{\gamma} + \beta_3 M^{\gamma} + \beta_4 S^{\gamma})^{\rho/\gamma}]^{1/\rho},$$

where α_i, $\beta_i > 0$, $\alpha_1 + \alpha_2 = \beta_1 + \ldots + \beta_4 = 1$, and ρ, $\gamma < 1$. The outer group, L and KEMS, has $\sigma_1 = |1-\rho|^{-1}$; and the inner group, K, E, M, and S, has $\sigma_2 = |1-\gamma|^{-1}$.

Figure 4 suggests a two-tiered TCES model with so-called L-E-KMS input groups. The top panel of figure 4 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 during 1949-1972 and 1982-2001 and rise sharply during 1973-1981. The bottom panel of figure 3 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 sharply until 1973 and thereafter grows slowly. In particular, the bottom panel of figure 4 suggests a two-tiered TCES model: an "outer" group of L, E, and KMS, with high substitution σ_1, and an "inner" group of K, M, and S, with low substitution σ_2. Because the bottom panel of figure 4 shows that K, M, and S move in close to fixed proportions, we expect σ_2 to be small. The relative constancy of L in figure 4 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 Solow residual. We may consider nonneutral technical change in the future. The L-E-KMS two-tiered CES model, $TCES_2$, takes the form

$$(3.2) \qquad Q = [\alpha_1 L^{\rho} + \alpha_2 E^{\rho} + \alpha_3 (\beta_1 K^{\gamma} + \beta_2 M^{\gamma} + \beta_3 S^{\gamma})^{\rho/\gamma}]^{1/\rho},$$

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

Tables 2 and 3 contain ICs from the $TCES_1$ and $TCES_2$ models. Because there are outer and inner elasticities of substitution in the TCES models, DF is equal to 47 in the constant-cost share models, is 43 in the IMA-cost share models, and remains 0 for the Törnqvist-cost share models. In the TCES models, the outer elasticity of substitution is $\sigma_1 \in \{.5, 1\}$ and the inner elasticity of substitution is $\sigma_2 \in \{.1, .17, .5, .67\}$. In the $TCES_1$ models, IMA-cost-share model 23, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs and positive DF. Similarly, in the $TCES_2$ models, IMA-cost-share model 35, with $\sigma_1 = 1$ and $\sigma_2 = .67$, has the lowest ICs and positive DF.

Among the 15 CES and 24 TCES models, the IMA-cost-share model 23 has the lowest ICs, and therefore, is the best model. The results reject a single elasticity of substitution for all the KLEMS inputs and suggest that TFP computed from the IMA-cost-share L-KEMS TCES model 23 is more appropriate than TFP computed from a Törnqvist-cost-share CD model.

The MSP method worked accurately for all models and sample periods. The accuracy of MSP computations is measured as the largest absolute residual of the computed first-order conditions (FOC). For the KLEMS inputs, there are six scalar FOC, five marginal productivity conditions and a cost line. Each scalar FOC may be written as a scalar expression equal to zero. The computed value of each scalar expression is an FOC residual, which we want to be as close to zero as possible. For each sample period, the MSP method computes the input residuals in many steps. For each step, the method computes six absolute FOC residuals. For each case, the optimal inputs, hence, the residual inputs, were computed so that the FOC were satisfied with approximately double-precision or 10^{-14} accuracy. Because economic data usually have no more than 5-6 decimal digits, 8 decimal digit accuracy exceeds the accuracy of usual economic data.

3.3. Optimal TFP versus Solow-Residual TFP.

We use the best production-function model, IMA-cost-share $TCES_1$ model 23, to compute $\%\Delta TFP_t^* = \%\Delta Q_t - \%\Delta f(\hat{v}_t)$, namely, period-to-period percentage change in optimal TFP, where $\%\Delta Q_t$ denotes period-to-period percentage change in observed output, $\%\Delta f(\hat{v}_t)$ denotes period-to-period percentage change in computed optimal output, for the best production function, $f(\cdot)$, at optimal inputs, $\hat{v}_t$. Similarly, let $\%\Delta TFP_t^{SR} = \%\Delta Q_t - c_{kt}\cdot\%\Delta K_t - \ldots - c_{St}\cdot\%\Delta S_t$, denote percentage change in Solow-residual TFP, where c_{kt}, ..., c_{st} denote Törnqvist input-cost shares and $\%\Delta K_t$, ..., $\%\Delta S_t$ denote percentage changes in the KLEMS inputs. To compare the percentage changes in optimal and Solow-residual TFP, we graph their difference in figure 5.

Figure 5 shows slight steady decline in the mean and more pronounced decline in the variance of $\%\Delta TFP_t^* - \%\Delta TFP_t^{SR}$ from 1949 to 2001, a pattern which reflects larger input residuals in early years of the sample. Table 4 contains statistics of $\%\Delta TFP_t^* - \%\Delta TFP_t^{SR}$, which show that optimal TFP grew over the period about .10% faster than Solow-residual TFP. Although the average difference between $\%\Delta TFP_t^*$ and $\%\Delta TFP_t^{SR}$ is small during the period, the maximum difference between these quantities is 1.91%.

4. Conclusion.

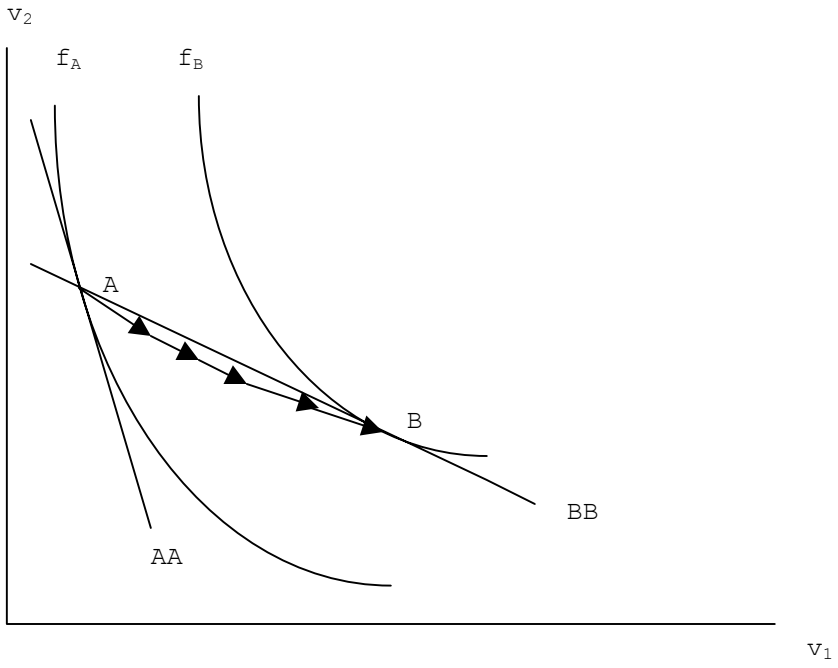
We have used the multi-step perturbation (MSP) method to compute input residuals for CES and TCES production functions of KLEMS inputs, for inelastic, unit elastic, and elastic input substitution, using KLEMS data from the Bureau of Labor Statistics, representing aggregate U.S. manufacturing from 1949-2001. We then used the input residuals to compute various ICs. We focus on ICs because, like log-likelihood functions, they provide a scalar measure of the empirical fit of a multiple equation model, in this case the five, numerically computed, demand functions of the KLEMS inputs. By extending $-(2/T) \times$ log-likelihood function with positive penalty terms, the ICs acknowledge that adding parameters is statistically costly. Being based on asymptotic arguments, ICs are not explicit about this, but fundamentally the problem is that, in the finite samples we are faced with in practice, adding parameters uses up degrees of freedom (DF). Adding too many parameters reduces DF to zero so that estimated parameters and derived quantities, such as TFP, based on them, have infinite

variances, hence, have no statistical reliability. BCAIC is the exception because it approaches $+\infty$ as DF approach zero, which is why BCAIC = $+\infty$ for models 11-15, 24-27, and 36-39, with DF = 0. Tables 1 to 3 present ICs ranging from unparsimonious AIC to parsimonious BIC.

For the results in tables 1 to 3, $TCES_1$ model 23, with IMA-cost shares, is the best model with the lowest ICs and positive DF. According to AIC, CD model 13 has a slightly lower AIC and, in this respect, is better but we dismiss Törnqvist models 11-15, 24-27, and 36-39 with DF = 0 as statistically unreliable. Figure 2 indicates that the IMA-cost shares and the Törnqvist-cost shares follow each other very closely. Thus, the Solow-residual TFP implied by model 13 should be close to TFP of model 23 computed as $\tau_t = q_t - f(\hat{v}_t)$. For the best TECS models, we graphed and tabulated summary statistics of the difference between the optimal TFP and Solow-residual TFP.

We searched for a best model for optimal TFP by choosing the elasticity of substitution over a coarse grid, conditional on estimated cost-shares. We shall consider using maximum likelihood to estimate all parameters jointly. Unless a measure of capacity is in the production function, joint estimation of production-function parameters often results in implausible estimates of input-cost-share parameters (Tatom, 1980). To avoid this problem, we would, first, use maximum likelihood to obtain initial elasticity estimates, for previously determined input-cost-share estimates, and would, then, use the initial parameter estimates as starting values in further joint maximum likelihood estimation of all parameters. Another extension would be to estimate more general production functions. In these extensions, we would use combined MSP and maximum likelihood to determine a best model which minimizes IC and compare its optimal TFP with Solow-residual TFP.

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.

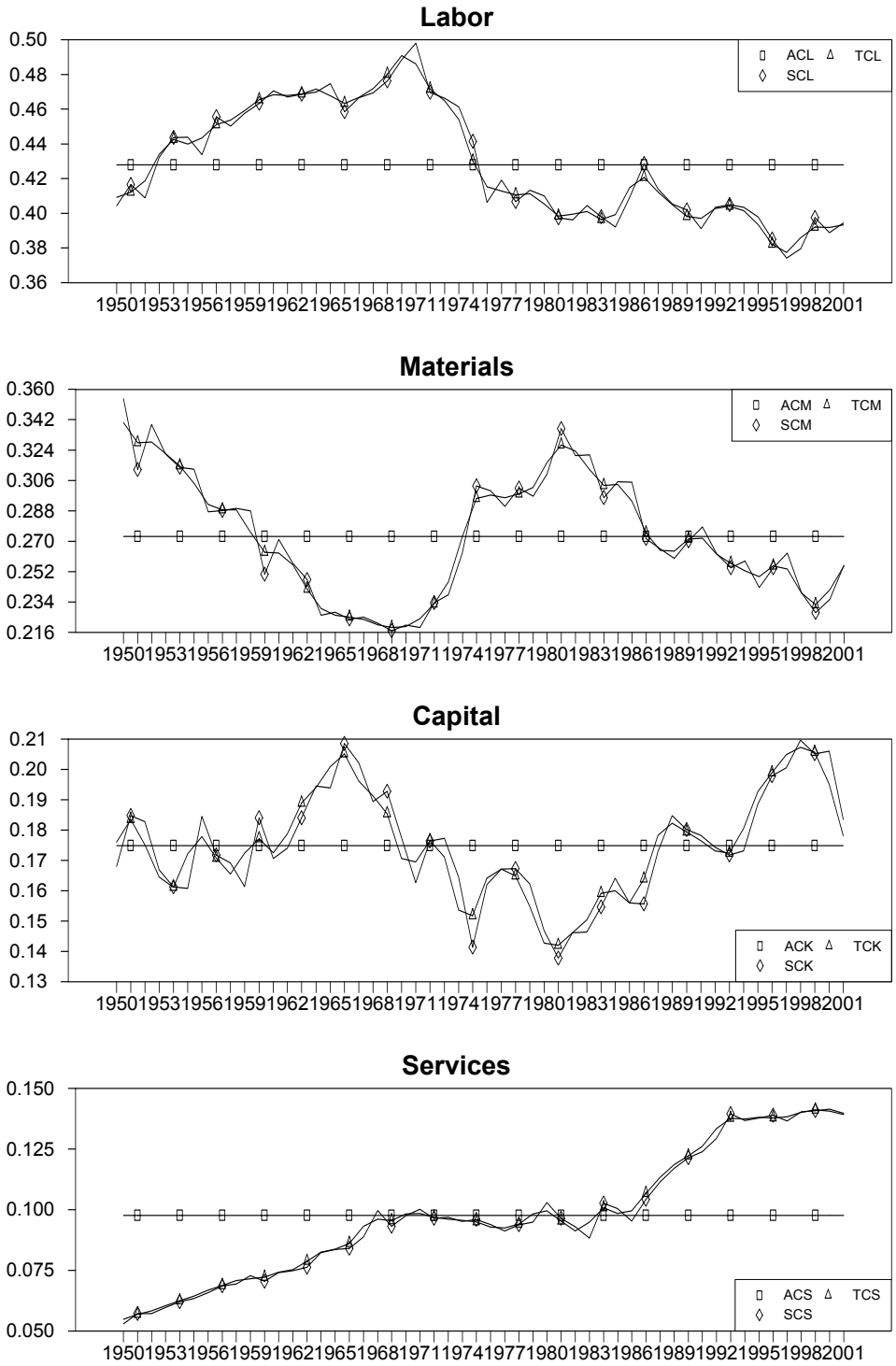


Figure 3: Scatter Plots of Pairwise Log of KLEMS Input Quantities.

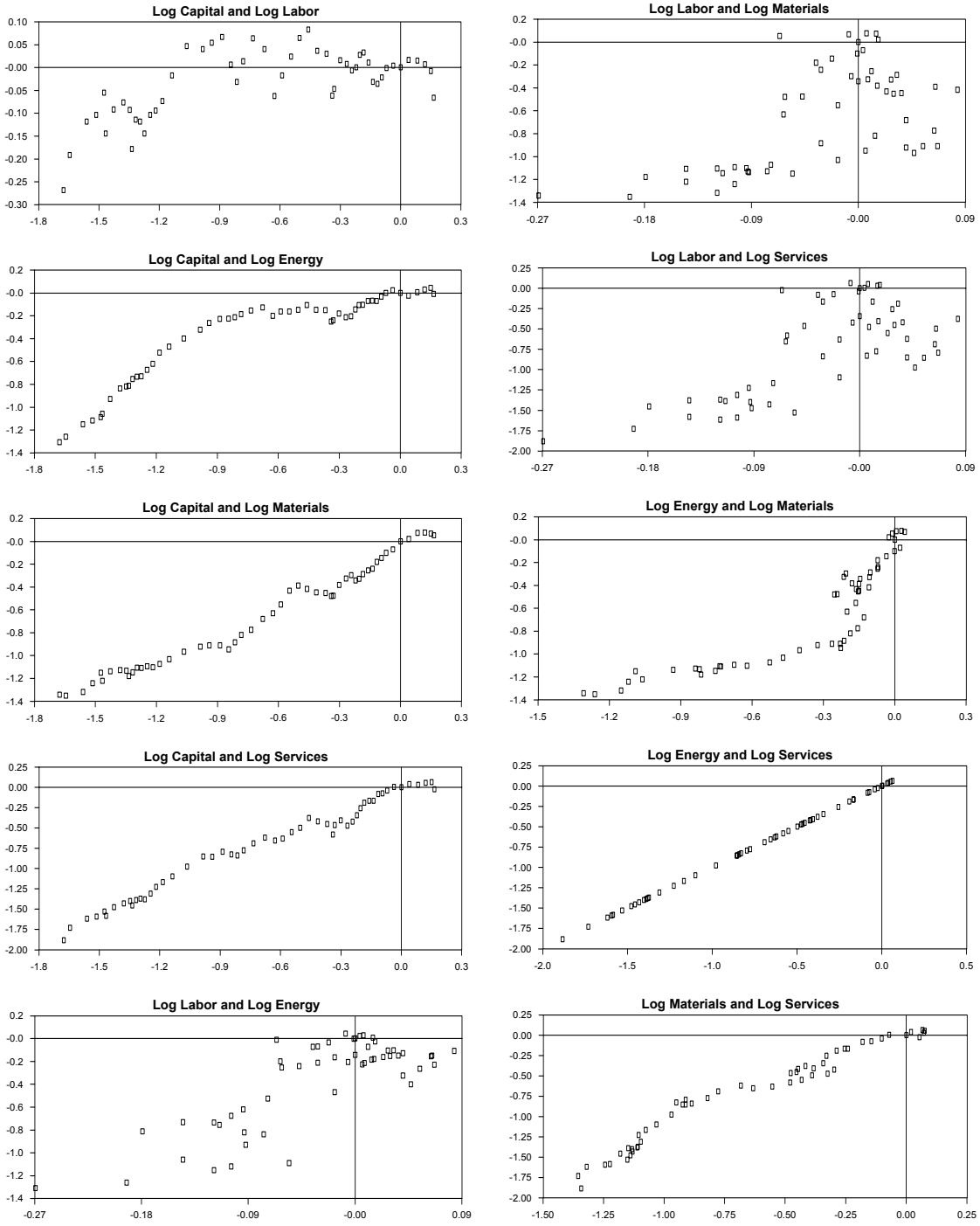
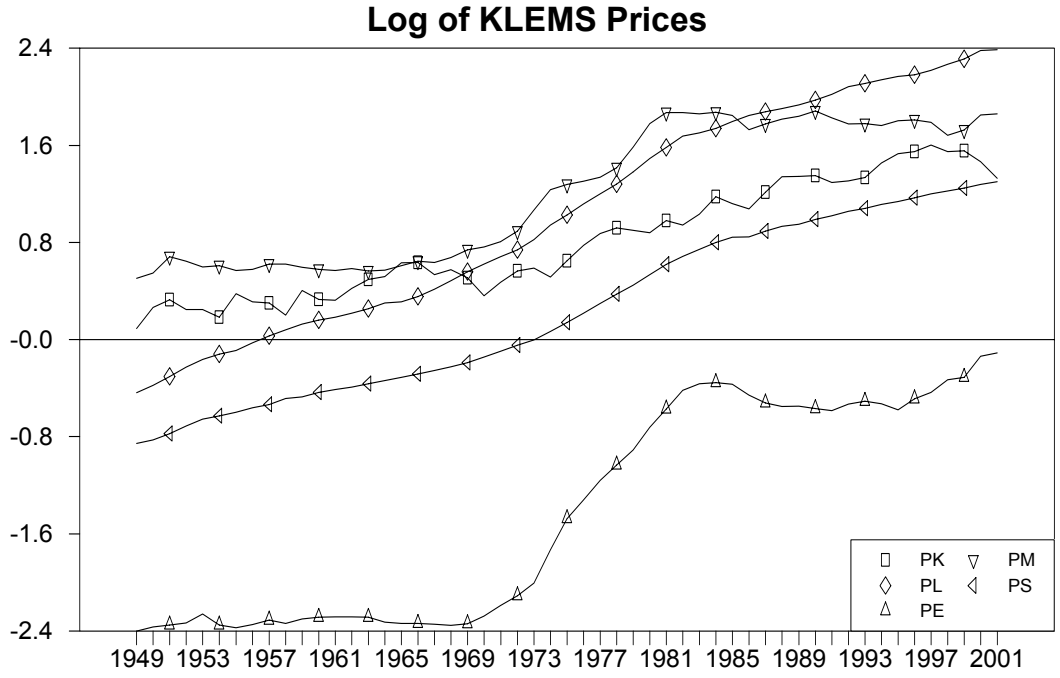


Figure 4: Log of KLEMS Input Prices and Quantities, 1949-2001.



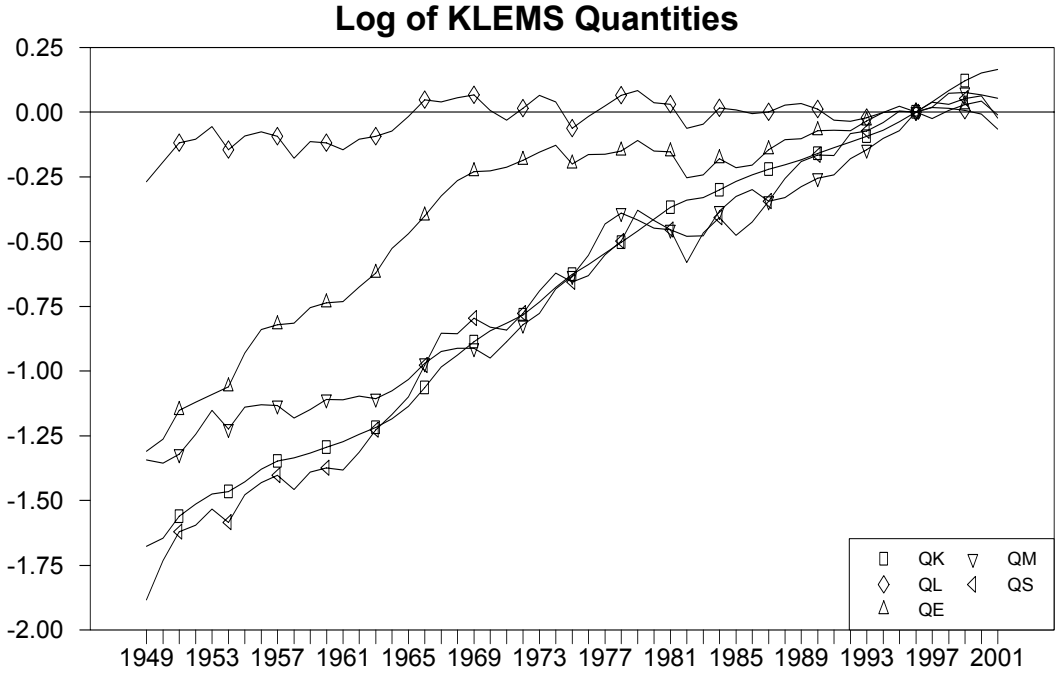


Figure 5: Graph of $\%\Delta TFP_t^* - \%\Delta TFP_t^{SR}$.

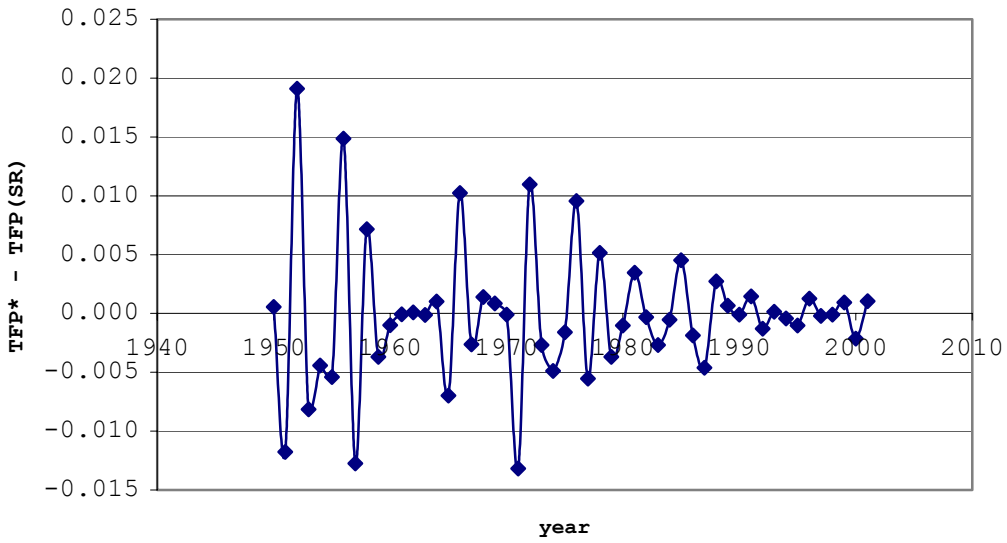


Table 1: Information Criteria of CES Models.

Model	Cost Shares	DF	σ	$-\left(\frac{2}{T}\right)\times LL$	AIC	BCAIC	BIC
1	Constant	48	.10	-25.37	-25.18	-25.16	-24.99
2	Constant	48	.50	-26.66	-26.47	-26.45	-26.29
3	Constant	48	1.0	-24.04	-23.86	-23.83	-23.67
4	Constant	48	2.0	-15.61	-15.42	-15.40	-15.24
5	Constant	48	10.	4.88	5.07	5.09	5.25
6	IMA	44	.10	-25.19	-24.85	-24.77	-24.52
7	IMA	44	.50	-27.21	-26.87	-26.79	-26.54
8	IMA	44	1.0	-32.62	-32.28	-32.21	-31.94
9	IMA	44	2.0	-16.34	-15.99	-15.92	-15.66
10	IMA	44	10.	3.92	4.26	4.34	4.59
11	TÖrnqvst	0	.10	-25.42	-17.39	$+\infty$	-9.47
12	TÖrnqvst	0	.50	-30.06	-22.02	$+\infty$	-14.10
13	TÖrnqvst	0	1.0	-42.22	-34.18	$+\infty$	-26.26
14	TÖrnqvst	0	2.0	-20.84	-12.80	$+\infty$	-4.88
15	TÖrnqvst	0	10.	.68	8.71	$+\infty$	18.63

Table 2: Information Criteria of $TCES_1$ Models.

No.	Cost Shares	DF	σ_1	σ_2	$-\left(\frac{2}{T}\right)LL$	AIC	BCAIC	BIC
16	Constant	48	.5	.1	-27.54	-27.36	-27.33	-27.17
17	Constant	48	.5	.17	-27.69	-27.50	-27.48	-27.32
18	Constant	48	.5	.5	-26.66	-26.47	-26.45	-26.29
19	Constant	48	.5	.67	-25.67	-25.48	-25.46	-25.30
20	IMA	44	1.	.1	-29.51	-29.17	-29.09	-28.84
21	IMA	44	1.	.17	-30.06	-29.72	-29.64	-29.38
22	IMA	44	1.	.5	-32.67	-32.33	-32.25	-32.00
23	IMA	44	1.	.67	-33.54	-33.21	-33.13	-32.87
24	TÖrnqvist	0	1.	.1	-30.88	-22.84	$+\infty$	-14.93
25	TÖrnqvist	0	1.	.17	-31.46	-23.42	$+\infty$	-15.50
26	TÖrnqvist	0	1.	.5	-35.01	-26.98	$+\infty$	-19.06
27	TÖrnqvist	0	1.	.67	-37.26	-26.98	$+\infty$	-19.06

Comment: The $TCES_1$ production-function model is $Q = [\alpha_1 L^{\rho} + \alpha_2(\beta_1 K^{\gamma} + \beta_2 E^{\gamma} + \beta_3 M^{\gamma} + \beta_4 S^{\gamma})^{\rho/\gamma}]^{1/\rho}$, for $\sigma_1 = |1-\rho|^{-1} \in \{.5, 1\}$ and $\sigma_2 = |1-\gamma|^{-1} \in \{.1, .17, .5, .67\}$.

Table 3: Information Criteria of $TCES_2$ Models.

No.	Cost Shares	DF	σ_1	σ_2	$-\left(\frac{2}{T}\right)\times LL$	AIC	BCAIC	BIC
28	Constant	48	.5	.1	-24.09	-23.90	-23.88	-23.71
29	Constant	48	.5	.17	-23.89	-23.70	-23.67	-23.51
30	Constant	48	.5	.5	-21.00	-20.81	-20.78	-20.62
31	Constant	48	.5	.67	-20.00	-19.81	-19.78	-19.62
32	IMA	44	1.	.1	-30.43	-30.09	-30.16	-29.76
33	IMA	44	1.	.17	-30.82	-30.48	-30.40	-30.15
34	IMA	44	1.	.5	-32.60	-32.26	-32.18	-31.92
35	IMA	44	1.	.67	-33.24	-32.90	-32.82	-32.56
36	TÖrnqvist	0	1.	.1	-35.88	-27.84	$+\infty$	-19.92
37	TÖrnqvist	0	1.	.17	-36.31	-28.27	$+\infty$	-20.35
38	TÖrnqvist	0	1.	.5	-38.89	-30.85	$+\infty$	-22.93
39	TÖrnqvist	0	1.	.67	-40.35	-32.32	$+\infty$	-24.40

Comment: The $TCES_2$ production-function model is $Q = [\alpha_1 L^{\rho} + \alpha_2 E^{\rho} + \alpha_3(\beta_1 K^{\gamma} + \beta_2 M^{\gamma} + \beta_3 S^{\gamma})^{\rho/\gamma}]^{1/\rho}$, for $\sigma_1 = |1-\rho|^{-1} \in \{.5, 1\}$ and $\sigma_2 = |1-\gamma|^{-1} \in \{.1, .17, .5, .67\}$.

Table 4: Statistics of $\%\Delta TFP^{*} - \%\Delta TFP^{SR}$.

$\%\Delta TFP^{*} - \%\Delta TFP^{SR}$	
Minimum	-1.32%
Maximum	1.91%
Mean	.10%
|Mean|	.39%
Std. Dev.	.60%

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.

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.

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

Ford, L.R. (1955), Differential Equations, New York, NY: McGraw-Hill.

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

Hannan, E.J. and B.G. Quinn (1979), "The Determination of the Order of an Autoregression," Journal of the Royal Statistical Society Series B 41: 190-195.

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

Judd, K.L. (1998), Numerical Methods in Economics, Cambridge, MA: M.I.T. Press.

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

Manzan, S. (2002), "Model Selection for Nonlinear Time Series," working paper, Department of Economics, University of Amsterdam, Amsterdam, The Netherlands.

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

Schwarz, G. (1978), "Estimating the Dimension 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.

Zadrozny, P.A. (2004), "Multi-Step Perturbation Method for Accurately Computing and Empirically Evaluating the Cost of Living," typescript, Bureau of Labor Statistics, Washington, DC.